Biological processes for the production of aryl sulfates.

Jendresen, Christian Bille; Nielsen, Alex Toftgaard

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 February 2016 (25.02.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/026976 A1

(51) International Patent Classification:
C12P 11/00 (2006.01) *C12N 9/10* (2006.01)

(21) International Application Number:
PCT/EP2015/069298

(22) International Filing Date:
21 August 2015 (21.08.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14182032.4 22 August 2014 (22.08.2014) EP

(71) Applicant: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 1, DK-2800 Kgs. Lyngby (DK).

(72) Inventors: **JENDRESEN, Christian Bille**; Haraldsgade 41, 4. th., DK-København N 2200 (DK). **NIELSEN, Alex Toftgaard**; Pennehave 3E, DK-2960 Rungsted Kyst (DK).

(74) Agent: **ZACCO DENMARK A/S**; Arne Jacobsens Allé 15, DK-2300 København S (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2016/026976 A1

(54) **Title**: BIOLOGICAL PROCESSES FOR THE PRODUCTION OF ARYL SULFATES

(57) **Abstract**: The present invention generally relates to the field of biotechnology as it applies to the production of aryl sulfates using polypeptides or recombinant cells comprising said polypeptides. More particularly, the present invention pertains to polypeptides having aryl sulfotransferase activity, recombinant host cells expressing same and processes for the production of aryl sulfates employing these polypeptides or recombinant host cells.

Biological processes for the production of aryl sulfates

Technical field of the invention

The present invention generally relates to the field of biotechnology as it applies to the production of aryl sulfates using polypeptides or recombinant cells comprising said polypeptides. More particularly, the present invention pertains to polypeptides having aryl sulfotransferase activity, recombinant host cells expressing same and processes for the production of aryl sulfates employing these polypeptides or recombinant host cells.

Background of the invention

A range of phenolic compounds are of great interest to the biotech industry since they are building blocks for polymeric compounds. Examples of such phenolic compounds include p-coumaric acid (pHCA) or other hydroxycinnamic acids which form the basis for many secondary metabolites including flavonoids and stilbenes. However, many of these phenolic compounds are toxic to producing organisms, and thus limit the productivity during fermentation. Hence, there is a need for large scale production processes, and especially for biological large scale production processes allowing improved productivity.

Moreover, a range of phenolic compounds, and especially those used as drugs or food additives such as resveratrol or vanillin, show poor solubility in water which makes it difficult for these compounds to be uptaken by the body. Hence, there is also a need for providing such phenolic compounds in a form which improves the solubility, and hence bioavailablility, preferably by using biological large scale production processes.

Summary of the invention

The object of the present invention is to provide a method for large scale production of aryl sulfates. Furthermore, it is an object to provide a biological process for the large scale production of phenols. The inventors have developed a biological process that solves both objects.

The present invention thus provides in a first aspect a process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

The present invention provides in a further aspect a process for the production of a sulfated phenolic compound, such as zosteric acid, the method comprises sulfating a phenolic compound, such as p-coumaric acid, using a polypeptide as detailed herein. Particularly, the process involves the use of a polypeptide having an aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g. SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

The present invention provides in a further aspect a recombinant host cell as detailed heren. Particularly, the present invention provides a recombinant host cell comprising (e.g. expresses) a first heterologous polypeptide having an aryl sulfotransferase activity, such as a heterologous polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

The present invention provides in a further aspect the use of a polypeptide as detailed herein in the sulfation of a phenolic compound. Particularly, the present invention provides the use of a polypeptide having an aryl sulfotransferase activity in the sulfation of a phenolic compound, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

The present invention provides in a further aspect a composition comprising a first recombinant host cell as detailed herein and a second recombinant host cell as detailed herein. Particularly, the present invention provides a composition comprising a first recombinant host cell comprising (e.g. expressing) a heterologous polypeptide having an

aryl sulfotransferase activity, such as a heterologous polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted; and

a second recombinant host cell comprising (e.g., expressing) a heterologous polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

The present invention provides in a further aspect a composition comprising a first polypeptide and a second polypeptide. Particularly, the present invention provides a

composition comprising a first polypeptide having an aryl sulfotransferase activity , such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted; and

a second polypeptide tyrosine ammonia lyase activity, such as a polypeptiode selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

Brief description of the drawings

Figure 1: Map of plasmid for expression of SULT1A1 from *Rattus norvegicus* in *Escherichia coli*

Figure 2: Map of plasmid for over-expression of cysDNC in *E. coli*.

Figure 3: Map of plasmid for over-expression of cysDNCQ in *E. coli*.

Figure 4: Map of plasmid for expression of RmXAL from *Rhodotorula mucilaginosa* / *Rhodotorula rubra* in *E. coli*.

Figure 5: Toxicity of unsulfated or sulfated products

Figure 6: Map of plasmid for expression of tyrosine ammonia-lyase RsTAL from *Rhodobacter sphaeroides* in *E.coli.*

Figure 7: Map of plasmid for expression of tyrosine ammonia-lyase FjTAL from *Flavobacterium johnsoniae* in *E.coli.*

Figure 8: Map of plasmid for expression of tyrosine ammonia-lyase RcTAL from *Rhodobacter capsulatus* in *E.coli*.

Figure 9: Map of plasmid for expression of SULT1A1 from *Rattus norvegicus* in *Saccharomyces cerevisiae* (native gene).

Figure 10: Map of plasmid for expression of SULT1A1 from *Rattus norvegicus* in *Saccharomyces cerevisiae* (codon-optimized gene).

Detailed description of the invention

Unless specifically defined herein, all technical and scientific terms used have the same meaning as commonly understood by a skilled artisan in the fields of biochemistry, genetics, and molecular biology.

All methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, with suitable methods and materials being described herein. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will prevail. Further, the materials, methods, and examples are illustrative only and are not intended to be limiting, unless otherwise specified.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (B. D. Harries & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In ENZYMOLOGY (J. Abelson and M. Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols.154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); and Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Polypeptides and host cells

As indicated above, the present invention provides and utilizes polypeptides having aryl sulfotransferase activity (EC:2.8.2.1). This makes them particularly suitable for the sulfation of phenolic compounds such as p-coumaric acid and derivatives thereof (e.g., caffeic acid, ferulic acid or sinapic acid), or resveratrol.

The polypeptide having aryl sulfotransferase activity may be a sulfotransferase 1A1 enzyme, a sulfotransferase 1A2 enzyme, a sulfotransferase 1A3 enzyme, a sulfotransferase 1B1 enzyme, a sulfotransferase 1C1 enzyme, a sulfotransferase 1C2 enzyme, a sulfotransferase 1C4 enzyme, or a sulfotransferase 1E1 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme. According to certain other embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1A2 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase

1B1 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C1 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C2 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C4 enzyme. According to other certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1E1 enzyme (estrogen sulfotransferase), such as the sulfotransferase 1E1 from *Gallus gallus domesticus*.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a mammalian aryl sulfotransferase, such as a mammalian sulfotransferase 1A1 enzyme.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is an aryl sulfotransferase from *Rattus norvegicus* or a variant thereof. Such variant may have at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence of the aryl sulfotransferase from *Rattus norvegicus*. Such variant may also have an amino acid sequence of the sulfotransferase from *Rattus norvegicus*, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference aryl sulfotransferase. The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

According to certain embodiemtns, the polypeptide having aryl sulfotransferase activity may be a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least

about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, a polypeptide according to the invention is a polypeptide according to a). Accordingly, a polypeptide according to the invention may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 1. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 2. According to yet other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 3. According to yet other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 4. According to yet other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 5. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 6. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 7. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 8. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 9. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 10. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 11. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 12. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 13.

According to other certain embodiments, a polypeptide according to the invention is a polypeptide according to b). Accordingly, a polypeptide according to the invention may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to more particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the

amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1.

Preferably, a polypeptide according to b) has aryl sulfotransferase activity. More preferably, a polypeptide according to b) has a aryl sulfotransferase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to certain embodiment, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 4. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 7. According to certain other

embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 9. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 10. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 11. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12. According to certain other embodiments, a polypeptide according to b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 13.

With "similar" aryl sulfotransferase activity, it is meant that the polypeptide according to b) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the aryl sulfotransferase activity of the reference polypeptide (e.g., SEQ ID NO: 1).

The aryl sulfotransferase activity may for instance be determined in accordance to the following method: Aryl sulfotransferase activity may be determined by the reaction of radioactively sulfur labelled PAPS, [^{35}S]PAPS, with the substrate in presence of the polypeptide of interest. This is described previously, for example by Hattori *et al* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). The reaction takes place in a buffer such as 250 µL 50 mM sodium phosphate pH 6.8 with 1 µM [^{35}S]PAPS (3.7kBq) with 100 µM accepting compound for a period of 30 min at 30°C. The reaction is stopped by addition of 100 µL of a 1:1 mixture of 0.1 M barium acetate and barium hydroxide. 50 µL of 0.1 M zinc sulfate is added, followed by centrifugation at 1,200 × g for 5 min. 300 µL of the supernatant is then transferred to a new container and 50 µL of an equal volume of 0.1 M barium hydroxide and 0.1 M zinc sulfate is added. The mixture is then centrifuged at 13,000 × g for 5 min, and

300-µL aliquots of the supernatant are mixed with 2.5 mL of Cleasol I (Nacalai Tesque, Kyoto, Japan). The radioactivity is then measured by scintillation.

Alterntively, the activity of a sulfotransferase may be detected by direct measurement of the product by analytical methods such as high performance liquid chromatography (HPLC) or liquid chromatography in combination with mass spectrometry (LC-MS).

According to other certain embodiments, a polypeptide according to the invention is a polypeptide according to c). Accordingly, a polypeptide according to the invention may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 or more, such as 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, 100 or more, 110 or more, 120 or more, 130 or more, 140 or more, or 150 or more, amino acid residues are substituted, deleted, and/or inserted. According to particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other more particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (e.g., SEQ ID NO: 1). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to c) has aryl sulfotransferase activity. More preferably, a polypeptide according to c) has a aryl sulfotransferase activity similar to that of the

polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to certain embodiment, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 4. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 7. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 9. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 10. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 11. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12. According to certain other embodiments, a polypeptide according to c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 13.

With "similar" aryl sulfotransferase activity it is meant that the polypeptide according to c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, at of the aryl sulfotransferase activity of the reference polypeptide (e.g., SEQ ID NO: 1).

The aryl sulfotransferase activity may for instance be determined in accordance to the following method: Aryl sulfotransferase activity may be determined by the reaction of radioactively sulfur labelled PAPS, [^{35}S]PAPS, with the substrate in presence of the polypeptide of interest. This is described previously, for example by Hattori *et al* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). The reaction takes place in a buffer such as 250 µL 50 mM sodium phosphate pH 6.8 with 1 µM [^{35}S]PAPS (3.7kBq) with 100 µM accepting compound for a period of 30 min at 30°C. The reaction is stopped by addition of 100 µL of a 1:1 mixture of 0.1 M barium acetate and barium hydroxide. 50 µL of 0.1 M zinc sulfate is added, followed by centrifugation at 1,200 × g for 5 min. 300 µL of the supernatant is then transferred to a new container and 50 µL of an equal volume of 0.1 M barium hydroxide and 0.1 M zinc sulfate is added. The mixture is then centrifuged at 13,000 × g for 5 min, and 300-µL aliquots of the supernatant are mixed with 2.5 mL of Cleasol I (Nacalai Tesque, Kyoto, Japan). The radioactivity is then measured by scintillation.

Alterntively, the activity of a sulfotransferase may be detected by direct measurement of the product by analytical methods such as high performance liquid chromatography (HPLC) or liquid chromatography in combination with mass spectrometry (LC-MS).

Contemplated by the present invention is the production of a sulfated phenolic compound from a precursor thereof, and in particular from a precursor of the general formula (p-I) as described in more detail below. In this case, it may be suitable to employ a further (e.g., second) polypeptide which has tyrosine ammonia lyase activity. Such polypeptide may be a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 or more, such as about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted, and/or inserted.

According to certain embodiments, a further polypeptide according to the invention is a polypeptide according to d). Accordingly, a polypeptide according to the invention may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14). According to particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 14. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 15. According to yet other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 16. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 17. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 18. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 19. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 20. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 21. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 22. According other particular embodiments, a polypeptide according to d) comprises an amino acid sequence set forth in SEQ ID NO: 23.

According to other certain embodiments, a further polypeptide according to the invention is a polypeptide according to e). Accordingly, a polypeptide according to the invention may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as

at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14). According to particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14). According to other particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14). According to other particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14). According to other particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14).

According to particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to more particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other more particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at

least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other more particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other more particular embodiments, a polypeptide according to e) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14.

Preferably, a polypeptide according to e) has tyrosine ammonia lyase activity. More preferably, a polypeptide according to e) has a tyrosine ammonia lyase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14).

According to certain embodiment, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 14. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 15. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 16. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 17. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 18. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 19. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 20. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 21. According to certain other embodiments, a

polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 22. According to certain other embodiments, a polypeptide according to e) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 23. With "similar" tyrosine ammonia lyase activity it is meant that the polypeptide according to e) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the ammonia lyase activity of the reference polypeptide (e.g., SEQ ID NO: 14).

The tyrosine ammonia lyase activity may for instance be determined in accordance to the following method: Enzymatic assays are performed in 200 µL volumes in wells in a UV transparent 96-well plate, by following the increase in absorbance at 315 nm (pHCA) using spectrophotometry or HPLC with UV detection. The reaction mixtures contain 2 µg of purified protein and are initiated by adding 1 mM tyrosine or 6 mM after equilibration to 30°C. The enzymatic activity is calculated as U/g, where U is defined as µmol substrate converted per minute. Negative controls contain no purified protein. Kinetic constants Km and vmax are determined from assays containing 1.56 µM to 200 µM tyrosine.

According to other certain embodiments, a further polypeptide according to the invention is a polypeptide according to f). Accordingly, a polypeptide according to the invention may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 or more, such as 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, 100 or more, 110 or more, 120 or more, 130 or more, 140 or more, or 150 or more, amino acid residues are substituted, deleted, and/or inserted. According to particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100,

about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 30, such as about 1 to about

25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to f) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (e.g., SEQ ID NO: 14). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to f) has tyrosine ammonia lyase activity. More preferably, a polypeptide according to f) has a tyrosine ammonia lyase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14). According to certain embodiment, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 14. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 15. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 16. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 17. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 18. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 19. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 20. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase

activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 21. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 22. According to certain other embodiments, a polypeptide according to f) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 23. With "similar" tyrosine ammonia lyase activity it is meant that the polypeptide according to f) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the ammonia lyase activity of the reference polypeptide (e.g., SEQ ID NO: 14).

The tyrosine ammonia lyase activity may for instance be determined in accordance to the following method: Enzymatic assays are performed in 200 µL volumes in wells in a UV transparent 96-well plate, by following the increase in absorbance at 315 nm (pHCA) using spectrophotometry or HPLC with UV detection. The reaction mixtures contain 2 µg of purified protein and are initiated by adding 1 mM tyrosine or 6 mM after equilibration to 30°C. The enzymatic activity is calculated as U/g, where U is defined as µmol substrate converted per minute. Negative controls contain no purified protein. Kinetic constants Km and vmax are determined from assays containing 1.56 µM to 200 µM tyrosine.

Further contemplated by the present invention is to employ a further (e.g., third) polypeptide which has phenylalanine ammonia lyase activity, such as a phenylalanine ammonia lyase (EC 4.3.1.24).

The polypeptides may be employed in accordance with the invention in isolated form, such as in purified form. The polypeptides may for instance be expressed by a recombinant host cell, and then purified. Techniques and means for the purification of polypeptides produced by a recombinant host cell are well know in the art. For example, in order to facilitate purification, an amino acid motif comprising several histidine residues, such as at least 6, may be inserted at the C- or N-terminal end of the polypeptide. A non-limiting example of such amino acid motif is provided in SEQ ID NO: 24. Various purification kits for histidine-

tagged polypeptides are available from commercial sources such as Qiagen, Hilden, Germany; Clontech, Mountain View, CA, USA; Bio-Rad, Hercules, CA, USA and others.

Alternatively, the polypeptides may be chemically synthezised. Techniques for chemical peptide synthesis are well know and include Liquid-phase synthesis and Solid-phase synthesis.

The polypeptides can also be employed in accordance with the invention as part of a recombinant host cell. Such recombinant host cells are described in more details below.

It is understood that the details given herein with respect to polypeptides apply to all aspects of the invention.

The present invention also provides recombinant host cells comprising (e.g. expressing) one or more polypeptides as detailed herein. Generally, the polypeptides according to the invention will be heterologous to the host cells, which means that the polypeptides are normally not found in or made (i.e. expressed) by the host cells, but derived from a different species.

Therefore, the present invention provides a recombinant host cell comprising a heterologous polypeptide having an aryl sulfotransferase activity. According to certain embodiments, a recombinant host cell according to the invention comprises a heterologous polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

A recombinant host cell provided and utilized in accordance with the present invention may comprise a hetereologus polypeptide having tyrosine ammonia lyase activity. According to certain embodiments, a recombinant host cell according to the invention comprises a heterologous polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 or more, such as about 1 to about 50, about 1 to about 40, about 1 to about 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, a recombinant host cell comprises a first heterologous polypeptide having aryl sulfotransferase activity and a second heterologous polypeptide having tyrosine ammonia lyase activity. According to particular embodiments, a recombinant host cell comprises a first heterologous polypeptide selected from the polypeptides according to items a) to c) as detailed herein, and a second heterologous polypeptide selected from the polypeptides according to items e) to f) as detailed herein.

According to more particular embodiments, a recombinant host cell comprises a first heterologous polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted; and

a second heterologous polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

Alternatively, the first heterologous polypeptide having aryl sulfotransferase activity may be comprised by a first recombinant host cell, and the second heterologous polypeptide having tyrosine ammonia lyase activity may be comprised by a second recombinant host cell.

According to certain embodiments, a recombinant host cell comprises a first heterologous polypeptide having aryl sulfotransferase activity and a further (e.g., third) heterologous polypeptide having phenylalanine ammonia lyase activity.

Alternatively, the first heterologous polypeptide having aryl sulfotransferase activity may be comprised by a first recombinant host cell, and the further (e.g., third) heterologous polypeptide having phenylalanine ammonia lyase activity may be comprised by a further (e.g., third) recombinant host cell. Such further recombinant host cell may be a recombinant host cell also comprising a heterologous polypeptide having tyrosine ammonia lyase activity.

Recombinant host cells in accordance with the invention can be produced from any suitable host organism, including single-celled or multicellular microorganisms such as bacteria, yeast, fungi, algae and plant, and higher eukaryotic organisms including nematodes, insects, reptiles, birds, amphibians and mammals.

According to certain embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, algae and plant.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, and algae.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast and fungi.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria and yeast.

According to certain embodiments, a recombinant host cells in accordance with the invention is not a plant cell.

Bacterial host cells are selected from Gram-positive and Gram-negative bacteria. Non-limiting examples for Gram-negative bacterial host cells include species from the genera Escherichia, Erwinia, Klebsiella and Citrobacter. Non-limiting examples of Gram-positive bacterial host cells include species from the genera Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Streptomyces, Streptococcus, and Cellulomonas.

According to certain embodiments, the recombinant host cell is a bacterium, which may be a bacterium of the genus Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus, or Yersinia.

According to particular embodiments, the recombinant host cell is a bacterium of the genus Bacillus. Non-limiting examples of a bacteria of the genus Bacillus are Bacillus subtitlis, Bacillus amyloliquefaciens, Bacillus licheniformis, and Bacillus mojavensis. According to more particular embodiments, the recombinant host cell is Bacillus subtitlis. According to other more particular embodiments, the recombinant host cell is Bacillus licheniformis.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Lactococcus. A non-limiting example of a bacterium of the genus Lactococcus is

Lactococcus lactis. According to more particular embodiments, the recombinant host cell is Lactococcus lactis.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Corynebacterium. A non-limiting example of a bacterium of the genus Corynebacterium is Corynebacterium glutamicum. According to more particular embodiments, the recombinant host cell is Corynebacterium glutamicum.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Streptomyces. A non-limiting examples of a bacterium of the genus Streptomyces are Streptomyces lividans, Streptomyces coelicolor, or Streptomyces griseus. According to more particular embodiments, the recombinant host cell is Streptomyces lividans. According to other more particular embodiments, the recombinant host cell is Streptomyces coelicolor. According to other more particular embodiments,, the recombinant host cell is Streptomyces griseus.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Pseudomonas. A non-limiting example of a bacterium of the genus Pseudomonas is Pseudomonas putida. According to more particular embodiments, the recombinant host cell is Pseudomonas putida.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Geobacillus. A non-limiting examples of a bacterium of the genus Geobacillus are Geobacillus thermoglucosidasius and Geobacillus stearothermophilus. According to more particular embodiments, the recombinant host cell is Geobacillus thermoglucosidasius. According to other more particular embodiments, the recombinant host cell is Geobacillus stearothermophilus.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Thermoanaerobacterium. A non-limiting example of a bacterium of the genus Pseudomonas is Thermoanaerobacterium thermosaccharolyticum. According to more particular embodiments, the recombinant host cell is Thermoanaerobacterium thermosaccharolyticum.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Escherichia. A non-limiting example of a bacterium of the genus Escherichia is

Escherichia coli. According to more particular embodiments, the recombinant host cell is Escherichia coli.

Yeast host cells may be derived from e.g., Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula, or Trichosporon.

According to certain embodiments, the recombinant host cell is a yeast, which may be a yeast is of the genus Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula, or Trichosporon.

According to particular embodiments, the recombinant host cell is a yeast of the genus Saccharomyces. A non-limiting example of a yeast of the genus Saccharomyces is Saccharomyces cerevisiae. According to more particular embodiments, the recombinant host cell is Saccharomyces cerevisiae.

According to particular embodiments, the recombinant host cell is a yeast of the genus Pichia. Non-limiting example of a yeast of the genus Pichia are Pichia pastoris and pichia kudriavzevii. According to more particular embodiments, the recombinant host cell is Pichia pastoris. According to other more particular embodiments, the recombinant host cell is pichia kudriavzevii.

Fungi host cells may be derived from, e.g., Aspergillus.

According to certain embodiments, the recombinant host cell is a fungus, such as a fungi of the genus Aspergillus. Non-limiting examples of a fungus of the genus Aspergillus are Aspergillus Oryzae, Aspergillus niger or Aspergillus awamsii. According to more particular embodiments, the recombinant host cell is Aspergillus Oryzae. According to other more particular embodiments, the recombinant host cell is Aspergillus niger. According to other more particular embodiments, the recombinant host cell is Aspergillus awamsii.

Algae host cells may be derived from, e.g., Chlamydomonas, Haematococcus, Phaedactylum, Volvox or Dunaliella.

According to certain embodiments, the recombinant host cell is an alga, which may be an algae of the genus Chlamydomonas, Haematococcus, Phaedactylum, Volvox or Dunaliella.

According to particular embodiments, the recombinant host cell is an alga cell of the genus Chlamydomonas. A non-limiting example of an alga of the genus Chlamydomonas is Chlamydomonas reinhardtii.

According to particular embodiments, the recombinant host cell is an alga cell of the genus Haematococcus. A non-limiting example of an alga of the genus Haematococcus is Haematococcus pluvialis.

According to other particular embodiments, the recombinant host cell is an alga cell of the genus Phaedactylum. A non-limiting example of an alga of the genus Phaedactylum is Phaedactylum tricornatum.

A plant host cell may be derived from, e.g., soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

According to certain embodiments, the recombinant host cell is a plant cell, such as a plant cell selected from the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

Generally, a recombinant host cell according to the invention has been genetically modified to express one or more polypeptides as detailed herein, which means that one or more exogenous nucleic acid molecules, such as DNA molecules, which comprise(s) a nucleotide sequence or nucleotide sequences encoding said polypeptide or polypeptides has been introduced in the host cell. Techniques for introducing exogenous nucleic acid molecule, such as a DNA molecule, into the various host cells are well-known to those of skill in the art, and include transformation (e.g., heat shock or natural transformation), transfection, conjugation, electroporation, microinjection and microparticle bombardment.

Accordingly, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide as detailed herein.

In order to facilitate expression of the polypeptide in the host cell, the exogenous nucleic acid molecule may comprise suitable regulatory elements such as a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said polypeptide.

Promoters useful in accordance with the invention are any known promoters that are functional in a given host cell to cause the production of an mRNA molecule. Many such promoters are known to the skilled person. Such promoters include promoters normally associated with other genes, and/or promoters isolated from any bacteria, yeast, fungi, alga or plant cell. The use of promoters for protein expression is generally known to those of skilled in the art of moleculer biology, for example, see Sambrook et al., Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. , 1989. The promoter employed may be inducible. The term "inducible" used in the context of a promoter means that the promoter only directs transcription of an operably linked nucleotide sequence if a stimulus is present, such as a change in temperature or the presence of a chemical substance ("chemical inducer"). As used herein, "chemical induction" according to the present invention refers to the physical application of a exogenous or endogenous substance (incl. macromolecules, e.g., proteins or nucleic acids) to a host cell. This has the effect of causing the target promoter present in the host cell to increase the rate of transcription. Alternatively, the promoter employed may be constitutive. The term "constitutive" used in the context of a promoter means that the promoter is capable of directing transcription of an operably linked nucleotide sequence in the absence of stimulus (such as heat shock, chemicals etc.).

Non-limiting examples of promoters functional in bacteria, such as Bacillus subtilis, Lactococcus lactis or Escherichia coli, include both constitutive and inducible promoters such as T7 promoter, the beta-lactamase and lactose promoter systems; alkaline phosphatase (phoA) promoter, a tryptophan (trp) promoter system, tetracycline promoter, lambda-phage promoter, ribosomal protein promoters; and hybrid promoters such as the tac promoter. Other bacterial and synthetic promoters are also suitable.

Non-limiting examples of promoters functional in yeast, such as Saccharomyces cerevisiae, include xylose promoter, GAL1 and GAL10 promoters, TEF1 promoter, and pgk1 promoter.

Non-limiting examples of promoters functional in fungi, such as Aspergillus Oryzae or Aspergillus niger, include promotors derived from the gene encoding Aspergillus oryzae TAKA amylase, Aspergillus niger neutral α-amylase, Aspergillus niger acid stable α-amylase, Aspergillus niger or Aspergillus awamsii glucoamylase (gluA), Aspergillus niger acetamidase, Aspergillus oryzae alkaline protease, Aspergillus oryzae triose phosphatase isomerase, Rhizopus meihei aspartic proteinase, and Rhizopus meihei lipase.

Non-limiting examples of promoters functional in alga, such as Haematococcus pluvialis, include the CaMV35S promoter, the SV40 promoter, and promoter of the Chlamydomonas reinhardtii RBCS2 gene and the promoter of the Volvox carteri ARS gene.

Non-limiting examples of promoters functional in plant cells include the Lactuca sative psbA promoter, the tabacco psbA promoter, the tobacco rrn16 PEP+NEP promoter, the CaMV 35S promoter, the 19S promoter, the tomate E8 promoter, the nos promoter, the Mac promoter, and the pet E promoter or the ACT1 promoter.

Besides a promoter, the exogenous nucleic acid molecule may further comprise at least one regulatory element selected from a 5' untranslated region (5'UTR) and 3' untranslated region (3' UTR). Many such 5' UTRs and 3' UTRs derived from prokaryotes and eukaryotes are well known to the skilled person. Such regulatory elements include 5' UTRs and 3' UTRs normally associated with other genes, and/or 5' UTRs and 3' UTRs isolated from any bacteria, yeast, fungi, alga or plant cell.

If the host cell is a prokaryotic organism, the 5' UTR usually contains a ribosome binding site (RBS), also known as the Shine Dalgarno sequence which is usually 3-10 base pairs upstream from the initiation codon. Meanwhile, if the host cell is an eukaryotic organism the 5' UTR usually contains the Kozak consensus sequence. An eukaryotic 5' UTR may also contain cis-acting regulatory elements.

The exogenous nucleic acid molecule may be a vector or part of a vector, such as an expression vector. Normally, such a vector remains extrachromosomal within the host cell which means that it is found outside of the nucleus or nucleoid region of the host cell.

According to certain embodiments, a recombinant host cell according to the invention does not express an endogenous PAPS-dependent aryl sulfotransferase.

It is also contemplated by the present invention that the exogenous nucleic acid molecule is stably integrated into the genome of the host cell. Means for stable integration into the genome of a host cell, e.g., by homologous recombination, are well known to the skilled person.

The sulfation reaction depends on the supply of sulfate from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) or transferred from another sulfated compound. The inventors have shown that the sulfation reaction can be enhanced by improving the supply of PAPS (3'-phosphoadenosine 5'-phosphosulfate) and, in addition, by the removal of the product 3'-phosphoadenosine 5'-phosphate (PAP). The improved supply is obtained by deregulation, mutation or overexpression of enzymes that increase PAPS concentration or similarly reduce PAP concentration. This is exemplified in Example 2, where an increased production of zosteric acid in *Escherichia coli* is obtained by increasing the expression of the genes cysD, cysN, and cysC which are responsible for production of PAPS. Without being bound to a specific theory, it is believed that an adenylyl moiety (AMP) of ATP is transferred to sulfate to form activated sulfate, or APS (adenosine 5'-phosphosulfate). This extremely unfavorable reaction is kinetically and energetically linked to the hydrolysis of GTP by the enzyme ATP sulfurylase, which is composed of two types of subunits: an adenylyl transferase (cysD) and a GTPase (cysN). APS is then phosphorylated at the 3'-hydroxyl to form PAPS (3'-phosphoadenosine 5'-phosphosulfate) in a reaction catalysed by APS kinase, which is encoded by cysC. Furthermore, the inventors have enhanced the production of zosteric acid even more by increasing the expression of the gene cysQ encoding a PAP phosphatase which is responsible for the removal of PAP.

Therefore, in order to further improve the production of a sulfated phenolic compound, such as zosteric acid, a recombinant host cell according to the present invention may be further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification; may be further modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification; and/or may be further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification. By "increased protein expression" it is meant that the amount of

the respective protein produced by the thus modified host cell is increased compared an identical host cell that does not carry said modification. More particularly, by "increase expression" it is meant that the amount of respective protein produced by the thus modified host cell is increased by at least 10% , such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700% at least 800%, at least about 900%, at least about 1000%, at least about 2000%, at least about 3000%, at least about 4000%, at least about 5000%, at least about 6000%, at least about 7000%, at least about 8000% at least about 9000% or at least about 10000%, compared an identical host cell that does not carry said modification. The amount of protein in a given cell can be determined by any suitable quantification technique known in the art, such as ELISA, Immunohistochemistry or Western Blotting.

According to certain embodiments, a recombinant host cell according to the invention has further been modified to have an increased protein expression an ATP sulfurylase compared to an identical host cell that does not carry said modification.

According to certain embodiments, a recombinant host cell according to the invention has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

According to certain embodiments, a recombinant host cell according to the invention has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

An increase in protein expression may be achieved by any suitable means well-know to those skilled in the art. For example, an increase in protein expression may be achieved by increasing the number of copies of the gene or genes encoding the respective protein (e.g., ATP sulfurylase, APS kinase and/or PAP phosphatase) in the host cell, such as by using (e.g., introducing into the host cell) a vectors comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule. An increase in protein expression may also be achieved by integration of at least a second copy of the gene or genes encoding the respective protein into the genome of the host cell. An increase in protein expression may also be achieved by increasing the strength of the promoter(s) operably linked to the gene or genes. An increase in protein expression may

also be achieved by modifying the ribosome binding site on the mRNA molecule encoding the respective protein (e.g., ATP sulfurylase, APS kinase and/or PAP phosphatase). By modifying the sequence of the ribosome binding site the translation initiation rate may be increased, thus increasing the translation efficiency.

ATP sulfurylase encoding genes for use according to the invention may for instance be the cysD and cysN genes from *Escherichia coli* (encoding SEQ ID NO: 25 and 26, respectively). Alternative ATP sulfurylase encoding genes include the Arabidopsis thaliana ATP sulfurylase ASAL gene (GenBank Accession No. U40715, Logan et al. (1996) J Biol Chem 271: 12227); the Allium cepa ATP-sulfurylase gene (Gen-Bank Accession No AF21154); the Lotus japonicus ATP sulfurylase gene (GenBank Accession No. AW164083); the Arabidopsis thaliana met3-1 ATP sulfurylase gene (Gen-Bank Accession No. X79210).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding a ATP sulfurylase.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 25 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25, provide that the sequence identity is not 100%, and a nucleotide sequence encoding iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26, provide that the sequence identity is not 100%. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide

sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 25 and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25, provide that the sequence identity is not 100%, and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26, provide that the sequence identity is not 100%. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25, provide that the sequence identity is not 100%, and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26, provide that the sequence identity is not 100%. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

An alternative ATP sulfurylase encoding gene for use according to the invention may for instance be the MET3 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 68).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide

sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 68 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 68. Preferably, the polypeptide according to ii) has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 68.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 68. Preferably, the polypeptide has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 68. Preferably, the polypeptide has ATP sulfurylase activity.

An alternative ATP sulfurylase encoding gene for use according to the invention may for instance be the ATP sulfurylase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 73).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 73 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%,

at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73. Preferably, the polypeptide according to ii) has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 73.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73. Preferably, the polypeptide has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 73. Preferably, the polypeptide has ATP sulfurylase activity.

In order to facilitate expression of the polypeptides in the host cell, the exogenous nucleic acid molecule may comprise suitable regulatory elements such as a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequences encoding said polypeptides.

An APS kinase encoding gene for use according to the invention may for instance be the cysC gene from *Escherichia coli* (encoding SEQ ID NO: 27).

In certain instances a single polypeptide has been shown to possess both an ATP sulfurylase and a 5'-adenylylsulfate kinase activity. For example, an ATP sulfurylase/APS kinase encoding gene has been isolated from mouse (GenBank Accession No. U34883, Li et al. (1995) J Biol Chem)70: 1945), and human (GenBank Accession No. AF033026, Yanagisawa

(1998) Biosci Biotechnol Biochem 62: 1037) sources. Other examples of such bifunctional enzyme include 3'-phosphoadenosine 5'-phosphosulfate synthase enzymes (PAPSS) from rat (Rattus norvegicus) (SEQ ID NO: 71 or 72).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an APS kinase.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27, provide that the sequence identity is not 100%. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27, provide that the sequence identity is not 100%. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set

forth in SEQ ID NO: 27, provide that the sequence identity is not 100%. Preferably, said polypeptide has APS kinase activity.

An alternative APS kinase encoding gene for use according to the invention may for instance be the MET14 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 69).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 69 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 69. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 69.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 69. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 69. Preferably, said polypeptide has APS kinase activity.

An alternative APS kinase encoding gene for use according to the invention may for instance be the APS kinase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 74).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 74 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 74. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 74.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 74. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 74. Preferably, said polypeptide has APS kinase activity.

Alternatively, a polypeptide having both an ATP sulfurylase and a APS kinase activity activity can be used, such as a 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an 3'-phosphoadenosine 5'-phosphosulfate synthase.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 71 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 71. Preferably, said polypeptide according to ii) has both an ATP sulfurylase and a APS kinase activity activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 71.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 71. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 71. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide

sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 72 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 72. Preferably, said polypeptide according to ii) has both an ATP sulfurylase and a APS kinase activity activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 72.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 72. Preferably, said polypeptide has both an ATP sulfurylase and APS kinase activity activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 72. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity activity.

In order to facilitate expression of the polypeptide in the host cell, the exogenous nucleic acid molecule may comprise suitable regulatory elements such as a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said polypeptide.

An PAP phosphatase encoding gene for use according to the invention may for instance be the cysQ gene from *Escherichia coli* (encoding SEQ ID NO: 28).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an PAP phosphatase.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, provide that the sequence identity is not 100%. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, provide that the sequence identity is not 100%. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, provide that the sequence identity is not 100%. Preferably, said polypeptide has PAP phosphatase activity.

An alternative PAP phosphatase encoding gene for use according to the invention may for instance be the MET22 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 70).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 70 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 70. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 70.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 70. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 70. Preferably, said polypeptide has PAP phosphatase activity.

An alternative PAP phosphatase encoding gene for use according to the invention may for instance be the PAP phosphatase encoding gene from *Bacillus subtilits* (encoding SEQ ID NO: 75).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 75 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 75. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 75.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 75. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 75. Preferably, said polypeptide has PAP phosphatase activity.

In order to facilitate expression of the polypeptide in the host cell, the exogenous nucleic acid molecule may comprise suitable regulatory elements such as a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said polypeptide.

It is understood that the details given herein with respect to a recombinant host cell apply to other aspects of the invention, in particular to the processes and uses according to the invention, which are described in more detail below.

Methods and uses

The present invention provides processes for the production of sulfated phenolic compounds. Particularly, a process for the production of a sulfated phenolic compound is provided comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

According to certain embodiments, the process for the production of a sulfated phenolic compound comprises:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

According to other certain embodiments, the process for the production of a sulfated phenolic compound comprises:

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

According to other certain embodiments, the process for the production of a sulfated phenolic compound comprises:

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

The medium employed may be any conventional medium suitable for culturing the host cell in question, and may be composed according to the principles of the prior art. The medium will usually contain all nutrients necessary for the growth and survival of the respective host cell, such as carbon and nitrogen sources and other inorganic salts. Suitable media, e.g. minimal or complex media, are available from commercial suppliers, or may be prepared according to published receipts, e.g. the American Type Culture Collection (ATCC) Catalogue of strains. Non-limiting standard medium well known to the skilled person include Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, MS broth, Yeast Peptone Dextrose, BMMY, GMMY, or Yeast Malt Extract (YM) broth, which are all commercially available. A non-limiting example of suitable media for culturing bacterial cells, such as B. subtilis, L. lactis or E. coli cells, including minimal media and rich media such as Luria Broth (LB), M9 media, M17 media, SA media, MOPS media, Terrific Broth, YT and others. Suitable media for culturing eukaryotic cells, such as yeast cells, are RPMI 1640, MEM, DMEM, all of which may be supplemented with serum and/or growth factors as required by the particular host cell being cultured. The medium for culturing eukaryotic cells may also be any kind of minimal media such as Yeast minimal media.

The fermentable carbon substrate may be any suitable carbon substrate know in the art, and in particularly any carbon substrate commonly used in the cultivation of microorganisms and/ or fermentation. Non-limiting examples of suitable fermentable carbon substates include carbohydrates (e.g., C5 sugars such as arabinose or xylose, or C6 sugars such as glucose), glycerol, glycerine, acetate, dihydroxyacetone, one-carbon source, methanol, methane, oils, animal fats, animal oils, plant oils, fatty acids, lipids, phospholipids, glycerolipids, monoglycerides, diglycerides, triglycerides, renewable carbon sources, polypeptides (e.g., a microbial or plant protein or peptide), yeast extract, component from a yeast extract, peptone, casaminoacids or any combination of two or more of the foregoing.

According to certain embodiments, the carbon substate is selected from the group consisting of C5 sugars (such as arabinose or xylose), C6 sugars (such as glucose or

fructose), lactose, sucrose, glycerol, glycerine, acetate, Corn steep liquor, yeast extract, component from a yeast extract, peptone, casaminoacids or combinations thereof.

According to certain embodiments, the medium comprises glucose.

According to certain other embodiments, the medium comprises glycerol.

According to certain other embodiments, the medium comprises acetate.

It is also contemplated to use starch as a carbon substrate. Depending on the microorganism used, the metabolization of starch may require the supplementation of beta-glucosidase, such as the beta-glucosidase from Neurospora crassa, to the medium. Alternatively, a recombination host cell according to the invention may be further genetically modified to express a beta-glucosidase, such as the beta-glucosidase from *Neurospora crassa*.

When a fermentable carbon substrate is employed it is thus possible that the recombinant host cell produces the phenolic compound or a precursor thereof directly from such primary carbon substrate.

According to certain embodiments, the process further comprises:

(ii) culturing the first recombinant host cell under suitable conditions for the production of the corresponding sulfated phenolic compound.

Suitable conditions for culturing the respective host cell are well known to the skilled person. Typically, the recombinant host cell is cultured at a temperature ranging from about 23 to about 60°C, such as from about 25 to about 40°C, such as at about 37°C. The pH of the medium may range from pH 1.0 to pH 14.0, such as from about pH 1 to about pH 2, from about pH 4 to about pH 11, from about pH 5 to about pH 10, from about pH 6 to about pH 10, or from about pH 7 to about pH 9.5, e.g. at pH 6.0, pH pH 7.0, pH. 7.5, pH 8.0, pH 8.5, pH 9.0, pH 9.5, pH 10.0, pH 10.5 or pH 11.0.

The process may further comprise iii) recovering the sulfated phenolic compound. The sulfated phenolic compound may be recovered by conventional method for isolation and purification chemical compounds from a medium. Well-known purification procedures include centrifugation or filtration, precipitation, and chromatographic methods such as e.g. ion exchange chromatography, gel filtration chromatography, etc.

Further provided is a process for the production of a sulfated phenolic compound, such as zosteric acid, the method comprises sulfating a phenolic compound, such as p-coumaric acid, using a polypeptide having aryl sulfotransferase activity as detailed herein. Such polypeptide may be selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

Suitable conditions for the sulfation reaction are well known to the skilled person. Typically, the sulfation reaction takes place at a temperature ranging from about 23 to about 60°C, such as from about 25 to about 40°C, such as at about 37°C. The deamination reaction may take place at a pH ranging from pH 1.0 to pH 14.0, such as from about pH 2 to about pH 11, such as from about pH 5 to about pH 10, from about pH 6 to about pH 10, or from about pH 7 to about pH 9.5, e.g. at pH 6.0, pH pH 7.0, pH. 7.5, pH 8.0, pH 8.5, pH 9.0, pH 9.5, pH 10.0, pH 10.5 or pH 11.0.

Also provide is the use of a polypeptide in the sulfation of a phenolic compound, said polypeptide having aryl sulfotransferase activity as detailed herein. Such polypeptide may be selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

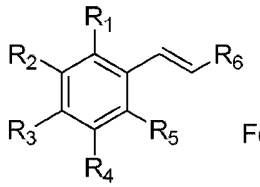
b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence

identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

For the purpose of this specification and the appended claims, it should be understood that the phenolic compounds include those compounds in which a hydroxyl group is directly attached to a benzenoid carbon atom, and which compounds may or may not contain other substituent groups.

According to certain embodiments, the phenolic compound is a compound represented by the general formula (I):



Formula (I)

;

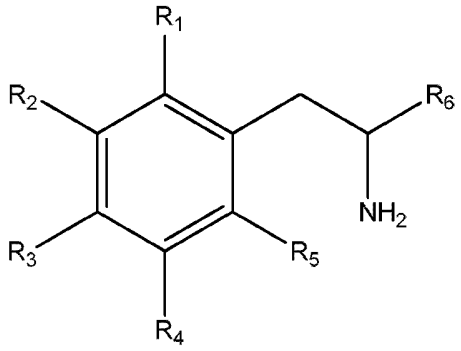
wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

Specific examples of compounds of Formula I include, but are not limited to, reservatrol, o-, m-, and p-coumaric acid, caffeic acid, ferulic acid, sinapic acid, curcumin, rosmarinic acid, sinapyl alcohol, coniferyl alcohol, and salvianolic acid.

A precursor of a phenolic compound according to Formula I may be a compound represented by the general Formula (p-I):



Formula (p-I);

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

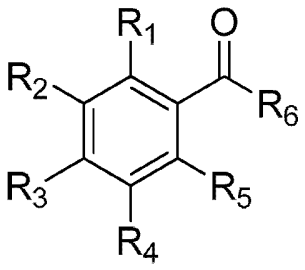
wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

Such a precursor may be converted to the phenolic compound by a recombinant host cell according to the invention, comprising a polypeptide having tyrosine ammonia lyase activity. Such polypeptide will eliminate ammonia from the precursor of Formula (p-I) under the formation of the corresponding molecule of Formula I. Preferably, the p-I precursor is the L-isomer.

Accoring to certain embodiments, the precursor of a phenolic compound as employed in step (i''') is a compound of the general Formula (p-I) as defined herein.

According to certain other embodiments, the phenolic compound is a compound represented by the general formula (II):



Formula (II)

;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

According to certain embodiments, R_6 is $-COOR_7$.

According to certain embodiments, R_7 is hydrogen.

According to certain embodiments, R_2 is hydroxyl (-OH).

According to certain embodiments, R_3 is hydroxyl (-OH).

According to certain embodiments, R_4 is hydroxyl (-OH).

According to certain embodiments, each of R_1, R_2, R_4 and R_5 is hydrogen.

According to certain embdodiments, each of R_1, R_2, and R_5 is hydrogen.

According to particular embodiments, the phenolic compound is p-coumaric acid (Formula I: R_1=H, R_2=H, R_3=OH, R_4=H, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic compound is caffeic acid (Formula I: R_1=H, R_2=H, R_3=OH, R_4=OH, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic acid is ferulic acid (Formula I: R_1=H, R_2=OCH_3, R_3=OH, R_4=H, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic acid is sinapic acid (Formula I: R_1=H, R_2=OCH_3, R_3=OH, R_4=OCH_3, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic compound is resveratrol (Formula I: R_1=H, R_2=OH, R_3=H, R_4=OH, R_5= H, R_6=p-hydroxyphenyl).

According to other particular embodiments, the phenolic compound is vanillin (Formula II: R_1=H, R_2=H, R_3=OH, R_4= OCH_3, R_5= H, R_6=H).

According to certain embodiments, the phenolic compound is a hydroxycinnamic acid.

According to certain embodiments, the phenolic compound is a compound represented by the general formula (I), wherein R_1 is hydrogen; R_2, R_3 and R_4 independently are selected from the group consisting of hydrogen (H), hydroxyl (-OH), C_{1-6}-alkyl and C_{1-6}-Alkoxy, provided that at least one of R_2, R_3 and R_4 is hydroxyl (-OH); R_5 is hydrogen, and R_6 is COOH.

According to certain embodiments, the precursor of a phenolic compound as employed in step (i''') is a compound of the general Formula (p-I), wherein R_1 is hydrogen; R_2, R_3 and R_4 independently are selected from the group consisting of hydrogen (H), hydroxyl (-OH), C_{1-6}-alkyl and C_{1-6}-Alkoxy, provided that at least one of R_2, R_3 and R_4 is hydroxyl (-OH); R_5 is hydrogen, and R_6 is COOH.

According to certain embodiment, the sulfated phenolic compound obtained in according to the present invention is zosteric acid.

Suitable sulfate donor molecules metabolized by a polypeptide having aryl sulfotransferase activity are well-known to one skilled in the art. Non-limiting examples include 3'-phosphoadenosine 5'-phosphosulfate (PAPS), para-nitrophenyl sulfate (pNPS) and 4-

methylumbelliferyl sulfate (MUS). Such sulfate donor molecules may be employed to facilitate the sulfation of phenolic compounds in accordance with the invention.

The medium employed for culturing the recombinant host cell may be any conventional medium suitable for culturing the host cell in question, and may be composed according to the principles of the prior art. The medium will usually contain all nutrients necessary for the growth and survival of the respective host cell, such as carbon and nitrogen sources and other inorganic salts, such as sulfate salts. Suitable media, e.g. minimal or complex media, are available from commercial suppliers, or may be prepared according to published receipts, e.g. the American Type Culture Collection (ATCC) Catalogue of strains. Non-limiting standard medium well known to the skilled person include Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, MS broth, Yeast Peptone Dextrose, BMMY, GMMY, or Yeast Malt Extract (YM) broth, which are all commercially available. A non-limiting example of suitable media for culturing bacterial cells, such as B. subtilis, L. lactis or E. coli cells, including minimal media and rich media such as Luria Broth (LB), M9 media, M17 media, SA media, MOPS media, Terrific Broth, YT and others. Suitable media for culturing eukaryotic cells, such as yeast cells, are RPMI 1640, MEM, DMEM, all of which may be supplemented with serum and/or growth factors as required by the particular host cell being cultured. The medium for culturing eukaryotic cells may also be any kind of minimal media such as Yeast minimal media.

Certain definitions

"Aryl sulfotransferase activity" as used herein refers to the ability of a polypeptide to catalyze the catalyze the transfer of a sulfate group from a donor molecule to an aryl acceptor molecule.

"Tyrosine ammonia lyase activity" as used herein refers to the ability of a polypeptide to catalysed the conversion of L-tyrosine into p-coumaric acid.

"Phenylalanine ammonia lyase activity" as used herein refers to the ability of a polypeptide to catalysed the conversion of L-phenylalanine into trans-cinnamic acid.

"ATP sulfurylase" as used herein refers to an enzyme that catalyzes the reaction: ATP + sulfate = diphosphate + adenosine 5'-phosphosulfate (APS).

“APS kinase” as used herein refers to an enzyme that catalyzes the reaction: ATP + adenosine 5’-phosphosulfate (APS) = ADP + 3’-phosphoadenosine 5’-phosphosulfate (PAPS).

“PAP phosphatase” as used herein refers to an enzyme that catalyzes the reaction: 3'-phosphoadenosine 5'-phosphate (PAP)+ H_2O = AMP + phosphate.

“Polypeptide," or "protein" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-transiational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

"Nucleic acid" or "polynucleotide" are used interchangeably herein to denote a polymer of at least two nucleic acid monomer units or bases (e.g., adenine, cytosine, guanine, thymine) covalently linked by a phosphodiester bond, regardless of length or base modification.

"Recombinant" or "non-naturally occurring" when used with reference to, e.g., a host cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant host cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

“Substitution” or “substituted” refers to modification of the polypeptide by replacing one amino acid residue with another, for instance the replacement of an Arginine residue with a Glutamine residue in a polypeptide sequence is an amino acid substitution.

"Conservative substitution" refers to a substitution of an amino acid residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain may be substituted with another aliphatic amino acid, e.g., alanine, valine, leucine, and isoleucine; an amino acid with hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain, e.g., serine and threonine; an amino acid having an aromatic

side chain is substituted with another amino acid having an aromatic side chain, e.g., phenylalanine, tyrosine, tryptophan, and histidine; an amino acid with a basic side chain is substituted with another amino acid with a basic side chain, e.g., lysine and arginine; an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain, e.g., aspartic acid or glutamic acid; and a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

"Non-conservative substitution" refers to substitution of an amino acid in a polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

"Deletion" or “deleted” refers to modification of the polypeptide by removal of one or more amino acids in the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide, in various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

"Insertion" or “inserted” refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. Insertions can comprise addition of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the reference polypeptide.

"Host cell" as used herein refers to a living cell or microorganism that is capable of reproducing its genetic material and along with it recombinant genetic material that has been introduced into it - e.g., via heterologous transformation.

"Expression" includes any step involved in the production of a polypeptide (e.g., encoded enzyme) including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

As used herein, "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded nucleic acid loop into which additional nucleic acid segments can be ligated. Certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". Certain other vectors are capable of facilitating the insertion of a exogenous nucleic acid molecule into a genome of a host cell. Such vectors are referred to herein as "transformation vectors". In general, vectors of utility in recombinant nucleic acid techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of a vector. Large numbers of suitable vectors are known to those of skill in the art and commercially available.

As used herein, "promoter" refers to a sequence of DNA, usually upstream (5') of the coding region of a structural gene, which controls the expression of the coding region by providing recognition and binding sites for RNA polymerase and other factors which may be required for initiation of transcription. The selection of the promoter will depend upon the nucleic acid sequence of interest. A "promoter functional in a host cell" refers to a "promoter" which is capable of supporting the initiation of transcription in said cell, causing the production of an mRNA molecule.

As used herein, "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence. A promoter sequence is "operably-linked" to a gene when it is in

sufficient proximity to the transcription start site of a gene to regulate transcription of the gene.

"Percentage of sequence identity," "% sequence identity" and "percent identity" are used herein to refer to comparisons between an amino acid sequence and a reference amino acid sequence. The “% sequence identify”, as used herein, is calculated from the two amino acid sequences as follows: The sequences are aligned using Version 9 of the Genetic Computing Group's GAP (global alignment program), using the default BLOSUM62 matrix (see below) with a gap open penalty of -12 (for the first null of a gap) and a gap extension penalty of -4 (for each additional null in the gap). After alignment, percentage identity is calculated by expressing the number of matches as a percentage of the number of amino acids in the reference amino acid sequence.

The following BLOSUM62 matrix is used:

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

"Reference sequence" or “reference amino acid sequence” refers to a defined sequence to which another sequence is compared. In the context of the present invention a reference amino acid sequence may be an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6,7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23.

Aliphatic radicals/groups, as referred herein, are optionally mono-or polysubstituted and may be branched or unbranched, saturated or unsaturated. Unsaturated aliphatic groups, as defined in herein, include alkyl, alkenyl and alkinyl radicals. Preferred aliphatic radicals according to the present invention include but are not restricted to methyl, ethyl, vinyl (ethenyl), ethinyl, propyl, n-propyl, isopropyl, allyl (2-propenyl), 1-propinyl, methylethyl, butyl, n-butyl, iso-butyl, sec-butyl, tert-butyl butenyl, butinyl, 1-methylpropyl, 2-methylpropyl, 1,1-dimethylethyl, pentyl, n-pentyl, 1,1-dimethylpropyl, 1,2-dimethylpropyl, 2,2-dimethylpropyl, hexyl, 1-methylpentyl, n-heptyl, n-octyl, n-nonyl and n-decyl. Preferred substituents for aliphatic radicals, according to the present invention, are a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, oxo, (C=O)R', SR', SOR', SO_2R', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents a linear or branched C_{1-6}-alkyl group.

“Alkyl”, “alkyl radical” or group as used herein means saturated, linear or branched hydrocarbons, which can be unsubstituted or mono- or polysubstituted. Thus, unsaturated alkyl is understood to encompass alkenyl and alkinyl groups, like e.g. -CH=CH-CH_3 or -C≡C-CH_3, while saturated alkyl encompasses e.g. -CH_3 and -CH_2-CH_3. “C_{1-12}-alkyl” includes C_{1-2}-alkyl, C_{1-3}-alkyl, C_{1-4}-alkyl, and C_{1-5}-alkyl, C_{1-6}-alkyl, C_{1-7}-alkyl, C_{1-8}-alkyl, C_{1-9}-alkyl, C_{1-10}-alkyl, and C_{1-11}-alkyl. In these radicals, C_{1-2}-alkyl represents C_1- or C_2-alkyl, C_{1-3}-alkyl represents C_1-, C_2- or C_3-alkyl, C_{1-4}-alkyl represents C_1-, C_2-, C_3- or C_4-alkyl, C_{1-5}-alkyl represents C_1-, C_2-, C_3-, C_4-, or C_5-alkyl, C_{1-6}-alkyl represents C_1-, C_2-, C_3-, C_4-, C_5- or C_6-alkyl etc. The alkyl radicals may be methyl, ethyl, vinyl (ethenyl), propyl, allyl (2-propenyl), 1-propinyl, methylethyl, butyl, 1-methylpropyl, 2-methylpropyl, 1,1-dimethylethyl, pentyl, 1,1-dimethylpropyl, 1,2-dimethylpropyl, 2,2-dimethylpropyl, hexyl, 1-methylpentyl, if substituted also CHF_2, CF_3 or CH_2OH etc. These alkyl, alkenyl or alkinyl radicals may optionally be mono-or polysubstituted by substitutents independently selected from a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, (C=O)R', SR', SOR', SO_2R', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents linear or branched C_{1-6}-alkyl group.

“Aryl” or “aryl radical” as herein is understood as meaning ring systems with at least one aromatic ring but without heteroatoms even in only one of the rings. These aryl radicals may optionally be mono-or polysubstituted by substitutents independently selected from a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, an optionally at least mono-

substituted phenyl group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, oxo, (C=O)R', SR', SOR', SO_2R', N(C=O)-OR', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents a linear or branched C_{1-6}-alkyl group. Preferred examples of aryl radicals include but are not restricted to phenyl, naphthyl, fluoranthenyl, fluorenyl, tetralinyl or indanyl or anthracenyl radicals, which may optionally be mono- or polysubstituted, if not defined otherwise.

"Alkyl-aryl" or "alkyl-aryl radical" as used herein comprises a linear or branched, optionally at least mono-substituted alkyl chain which is bonded to an aryl group, as defined above. A preferred alkyl-aryl radical is a benzyl group, wherein the alkyl chain is optionally branched or substituted. Preferred substituents for alky-aryl radicals, according to the present invention, are F, Cl, Br, I, NH_2, SH, OH, SO_2, CF_3, carboxy, amido, cyano, carbamyl, nitro, phenyl, benzyl, $-SO_2NH_2$, C_{1-6} alkyl and/or C_{1-6}-alkoxy.

"Heteroaryl" or "heteroaryl radical" as used herein is understood as meaning heterocyclic ring systems which have at least one aromatic ring and may optionally contain one or more heteroatoms from the group consisting of nitrogen, oxygen and/or sulfur and may optionally be mono-or polysubstituted by substitutents independently selected from a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, oxo, (C=O)R', SR', SOR', SO_2R', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents a linear or branched C_1-6-alkyl group. Preferred examples of heteroaryls include but are not restricted to furan, benzofuran, thiophene, benzothiophene, pyrrole, pyridine, pyrimidine, pyridazine, pyrazine, quinoline, isoquinoline, phthalazine, benzo-1,2,5-thiadiazole, benzothiazole, indole, benzotriazole, benzodioxolane, benzodioxane, benzimidzole, carbazole and quinazoline.

"Alkoxy", "alkoxy radical" or group as used herein means an "alkyl" singular bonded to oxygen. "C_{1-6}-alkoxy" includes C_{1-2}- alkoxy, C_{1-3}-alkoxy, C_{1-4}-alkoxy, and C_{1-5}-alkoxy, as well as C_{2-3}-alkoxy, C_{2-4}-alkoxy, C_{2-5}-alkoxy, C_{3-4}-alkoxy, C_{3-5}-alkoxy, and C_{4-5}-alkoxy. In these radicals, C_{1-2}- alkoxy represents C1- or C2- alkoxy, C_{1-3}- alkoxy represents C_1-, C_2- or C_3-alkoxy, C_{1-4}-alkyl represents C_1-, C_2-, C_3- or C_4-alkoxy, C_{1-5}-alkoxy represents C_1-, C_2-, C_3-, C_4-, or C_5-alkoxy, C_{1-6}- alkoxy represents C_1-, C_2-, C_3-, C_4-, C_5- or C_6-alkoxy. The alkoxy radicals may be methoxy, ethoxy, propoxy, butoxy, pentyloxy or hexyloxy.

The term "precursor of a phenolic compound" refers to any compound that may be converted to a phenolic compound by a host cells as described herein.

Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and sub ranges within a numerical limit or range are specifically included as if explicitly written out.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples, which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

Examples

Example 1 – Production of zosteric acid in E. coli

A range of aryl sulfotransferases including SULT1A1 *Rattus norvegicus* (SEQ ID NO: 1), SULT1A1 *Homo sapiens* (SEQ ID NO: 2), SULT1A1 *Equus caballus* (SEQ ID NO: 3), SULT1A1 *Sus scrofa domesticus* (SEQ ID NO: 4), SULT1A1 Canis lupus familiaris (SEQ ID NO: 5) and SULT1E1 *Gallus gallus domesticus* (SEQ ID NO: 6) were expressed in *Escherichia coli*. The respective genes encoding SEQ ID NO. 1, 3, 4, 5, and 6 were cloned amplified from liver tissue cDNA (Zyagen) by PCR using the primers listed in Table 1. The nucleotide sequence of the gene encoding SEQ ID NO: 2 was codon optimized for expression in *Escherichia coli* (GeneArt, Life Technologies) and amplified by PCR using the primers in Table 1. The pETDuet-1 plasmid was digested with restriction endonucleases *Nco*I and *Sal*I. The PCR products were then individually cloned into the plasmid pETDuet-1 using the Gibson reaction (New England Biolabs). The resulting plasmids were transformed into BL21(DE3)pLysS (Life Technologies). Figure 1 shows the plasmid map of the plasmid encoding SULT1A1 *Rattus norvegicus* (SEQ ID NO: 1).

Table 1: Overview of enzymes and primers for cloning aryl sulfotransferases

SEQ ID NO	Name	Fwd Primer	Rev Primer
1	SULT1A1 Rattus norvegicus	CBJP472	CBJP473
2	SULT1A1 Homo sapiens	CBJP470	CBJP471
3	SULT1A1 Equus caballus	CBJP499	CBJP500
4	SULT1A1 Sus scrofa domesticus	CBJP505	CBJP506
5	SULT1A1 Canis lupus familiaris	CBJP503	CBJP504
6	SULT1E1 Gallus gallus domesticus	CBJP501	CBJP502

The strains were grown in M9 minimal media containing glucose as a carbon source, and 0.1 mM IPTG for induction of gene expression as well as 0.1 mM *p*-coumaric acid (pHCA). After four days of growth, samples were withdrawn by filtration and analyzed by HPLC.

The concentration of p-courmaric acid (pHCA) and zosteric acid in the supernatant was quantified by high performance (HPLC) and compared to chemical standards. HPLC was done on a Thermo setup using a HS-F5 column and mobile phases: 5 mM ammonium formate pH 4.0 (A) and acetonitrile (B) at 1.5 mL min-1, using a gradient elution starting at 5% B. From 0.5 min after injection to 7 min, the fraction of B increased linearly from 5% to 60%, and between 9.5 min and 9.6 the fraction of B decreased back to 5%, and remaining there until 12 min. pHCA and zosteric acid were quantified by measuring absorbance at 277 nm.

Table 2 shows the remaining pHCA and the produced zosteric acid in the culture media. Zosteric acid was formed with an aryl sulfotransferase heterologously expressed in a microorganism exemplified by *E. coli* supplied with the substrate.

Table 2: Production of zosteric acid in E. coli from pHCA through the heterologous expression of sulfotransferases.

Enzyme	pHCA remaining (mM)	Zosteric acid formed (mM)
No enzyme	0.10	Not detectable
SULT1A1 Rattus norvegicus	0.02	0.10
SULT1A1 Homo sapiens	0.08	0.02
SULT1A1 Equus caballus	0.09	0.01
SULT1A1 Sus scrofa domesticus	0.09	0.01
SULT1A1 Canis lupus familiaris	0.10	0.01
SULT1E1 Gallus gallus domesticus	0.08	0.01

Example 2 – Increased production of zosteric acid in *E. coli*

The addition of sulfated groups to targets is **dependent** on supply of the donor molecule 3'-Phosphoadenosine 5'-phosphosulfate (PAPS). We examined if we could increase the production of zosteric acid by overexpressing enzymes providing PAPS and an enzyme that removes the product 3'-Phosphoadenosine 5'-phosphate (PAP).

Table 3: Cloning of enzymes involved in activating sulfate and product removal.

Genes	Fwd Primer	Rev Primer
cysDNC alone	CBJP491	CBJP492
cysDNC for artificial operon	CBJP491	CBJP497
cysQ for artificial operon	CBJP498	CBJP496

In E. coli, the genes cysD and cysN encode the two subunits of ATP sulfurylase (EC:2.7.7.4), cysC encodes APS kinase (EC:2.7.1.25), and cysQ encode a PAP phosphatase.

The cysDNC cluster was amplified by PCR from *E. coli* MG1655 chromosomal DNA using the primers shown in table3. The plasmid pRSFDuet-1 (Life Technologies) was digested by the restriction endonucleases NdeI and BglII. The gene cluster was inserted into the digested plasmid using the Gibson reaction (New England Biolabs). Figure 2 shows the resulting plasmid. For the combined expression of cysDNC and cysQ in an artificial operon, cysDNCQ, the two parts were amplified by PCR from *E. coli* MG1655 chromosomal DNA using the primers shown in Table 3. Again the parts were inserted into the digested plasmid. Figure 3 shows the resulting plasmids. The plasmid expressing SULT1A1 Homo sapiens (SEQ ID NO: 2) from example 1 was co-transformed into *E. coli* BL21(DE3)pLysS cells (Life Technologies) with either the plasmid expressing cysDNC or cysDNCQ.

Cells were grown as in Example 1 and the supernatants were analyzed for product formation as in example 1. The strain expressing SULT1A1 in combination with cysDNCQ was also grown without the addition of IPTIG for induction. Table 4 shows the concentrations of pHCA and zosteric acid.

Table 4: Concentrations of pHCA and zosteric acid in culture media with E. coli expressing an aryl sulfotransferase in combination with cysDNC and cysQ.

Enzymes	Induction	pHCA remaining (mM)	Zosteric acid formed (mM)
SULT1A1 Homo sapiens	0.1 mM IPTG	0.08	0.02
SULT1A1 Homo sapiens, CysDNC	0.1 mM IPTG	0.06	0.06
SULT1A1 Homo sapiens, CysDNCQ	0.1 mM IPTG	0.04	0.09
SULT1A1 Homo sapiens, CysDNCQ	None	0.10	Not detectable

This shows that more of the pHCA is transformed into zosteric acid when the protein expression of cysDNC is increased. Even more zosteric acid is formed when the protein expression cysQ is additionally increased.

Example 3 – A sulfated product can be formed in vivo by co-expression of an heterologous pathway and an aryl sulfotransferase

The production of a sulfated product can be accomplished biologically by the expression of aryl sulfotransferase as shown in example 1. The substrate for sulfation may also be formed by a biological organism, and here it will be shown for an organism expressing both a heterologous pathway leading to a phenolic compound and expressing a sulfotransferase acting upon the phenolic compound.

The enzyme RmXAL from Rhodotorula mucilaginosa / Rhodotorula rubra (SEQ ID NO: 20) has tyrosine ammonia lyase activity, thus catalyzing the non-oxidative deamination of the amino acid tyrosine, releasing p-coumaric acid (pHCA) and ammonia. The gene encoding RmXAL was codon optimized using standard algorithms for expression in *E. coli* available by GeneArt (Life Technologies) and amplified by PCR using the primers shown in table 5 and inserted into the pCDFDuet-1 vector (Novagen / Life Technologies), which had been digested by the restriction enzymes NdeI and BglII, using Gibson reaction (New England Biolabs). Figure 4 provides an image of the plasmid expressing RmXAL.

Table 5: Primers used for cloning of tyrosine ammonialyase

Genes	Fwd Primer	Rev Primer
RmXAL	CBJP487	CBJP488

The resulting plasmid was co-transformed into *E. coli* BL21(DE3)pLysS cells (Life Technologies) alone or together with the plasmid expressing SULT1A1 from Homo sapiens (example 1). The resulting strains was grown in M9 media with glucose as a carbon source, with 0.1 mM IPTG for induction of gene expression. Samples were taken as described previously (example 1) for analysis of product formation. Table 6 shows the resulting concentrations of pHCA and zosteric acid. RmXAL allowed the production of pHCA without addition of any substrate, thus providing a heterologous pathway from the cells normal metabolism to a heterologous product. The additional expression of an aryl sulfotransferase, exemplified by SULT1A1 from *Homo sapiens*, allowed the *in vivo*

conversion of pHCA to zosteric acid. Thus, an aryl sulfotransferase can act upon a compound produced *in vivo* and the cells can release the resulting sulfated product to the medium.

Table 6: Concentrations of pHCA and zosteric acid in culture media with E. coli expressing an aryl sulfotransferase in combination with a tyrosine ammonia lyase.

Enzymes	pHCA (mM)	Zosteric acid formed (mM)
RmXAL	0.04	Not detectable
SULT1A1 Homo sapiens, RmXAL	0.02	0.01

Example 4 – Decreased toxicity of sulfated product

E. coli MG1655 was grown in chemically defined M9 minimal media with 0.2% glucose as a carbon source without further addition or with the additions of either 10 mM, 20 mM, 25 mM, 30 mM, 35 mM or 40 mM p-coumaric acid (pHCA), or with 20 mM or 40 mM of the sulfate ester of pHCA (zosteric acid). All media preparations had been adjusted to pH 7. Cells were grown at 37°C with 250 rpm shaking in an orbital shaker. The growth rates were examined by following the optical density at 600 nm. The resulting growth rates in exponential growth phase are shown in Figure 5. Filled squares represent growth rates in media with pHCA. Open squares represent growth rates in media with zosteric acid. And the circle represents the growth rate in media without any of these additions. It is evident that the presence of pHCA is toxic to the cells, while the sulfate ester, zosteric acid is much less so.

Example 5 – In vivo supply of precursor of sulfated product

The substrate that is the subject for sulfation may be supplied to the medium void of such precursors or may be provided by microorganisms in the medium. Here we show that *p*-coumaric acid that is sulfated to generate zosteric acid, can be produced in vivo by the expression of a tyrosine ammonia-lyase.

The genes encoding the tyrosine ammonia-lyases RcTAL (from *Rhodobacter capsulatus*; SEQ ID NO: 48), RsTAL (from *Rhodobacter sphaeroides*; SEQ ID NO: 17) and FjTAL (from *Flavobacterium johnsoniae*; SEQ ID NO: 14) were cloned into expression vectors as follows. Genes (SEQ ID NO: 49, 50, and 51, respectively) were optimized for *E. coli* and synthesized by GeneArt, amplified by PCR using the oligonucleotides shown in the table below, and

cloned into pCDFDuet-1 (Novagen): The plasmid was digested with *Nde*I and *Bgl*II and gel purified. The genes were inserted by isothermal assembly using Gibson Assembly Master Mix (New England Biolabs), and transformed into chemically competent DH5α (laboratory strain) or NEB5α (New England Biolabs), selecting for resistance to 50 µg mL^{-1} spectinomycin in LB medium. Resulting plasmids pCBJ215 (RsTAL), pCBJ228 (FjTAL) and pCBJ297 (RcTAL) (Figures 6 to 8, respectively) were co-transformed by electroporation into the *E. coli* expression strain BL21(DE3) (Invitrogen/Life Technologies) together with a pETDuet-1-based plasmid expressing SULT1A1 from rat (Example 1). Transformation cultures were plated on LB containing 50 µg mL^{-1} spectinomycin and 100 µg mL^{-1} ampicillin. A control strain carrying pCDFDuet-1 was also made.

Primers:

Oligonucleotide	Gene	Direction	Sequence
CBJP483	RsTAL	Forward	CATCTTAGTATATTAGTTAAGTATAAGAAGGAG ATATACATATGCTGGCAATGAGCCCT
CBJP484	RsTAL	Reverse	TGGCCGGCCGATATCCAATTGATTAAACCGGAC TCTGTTGC
CBJP555	FjTAL	Forward	CATCTTAGTATATTAGTTAAGTATAAGAAGGAG ATATACATATGAACACCATCAACGAATATCTG
CBJP556	FjTAL	Reverse	TGGCCGGCCGATATCCAATTGATTAATTGTTAA TCAGGTGGTCTTTTACTTTCTG
CBJP745	RcTAL	Forward	CATCTTAGTATATTAGTTAAGTATAAGAAGGAG ATATACATATGCTGGATGCAACCATTGG
CBJP746	RcTAL	Reverse	TGGCCGGCCGATATCCAATTGATTATGCCGGA GGATCCGCT

Strains harboring recombinant plasmids were pre-cultured in 2xYT liquid medium with 100 µg mL^{-1} ampicillin and 50 µg mL^{-1} spectinomycin and incubated at 37°C and 250 rpm overnight. The following day, each pre-culture was transferred into 5 ml of M9 minimal medium with 0.2% glucose, 2 mM tyrosine and 1 mM IPTG for induction of expression. Cultures were placed in an incubator at 37°C with shaking at 250 rpm overnight. The supernatants were then collected by centrifugation twice and applied to HPLC analysis as described in example 1, and the titers of *p*-coumaric acid (pHCA) and zosteric acid (ZA) were quantified using chemical standards and are presented in the table below.

Sulfotransferase	Tyrosine ammonia-lyase	µM pHCA	µM ZA
SULT1A1 rat	None	0	0
SULT1A1 rat	RsTAL	78	<1
SULT1A1 rat	RcTAL	20	<1
SULT1A1 rat	FjTAL	398	16

Here, it is evident that the zosteric acid is formed when there is a supply of exogenous *p*-coumaric acid or if the cells are capable of producing *p*-coumaric acid. Conclusively, a sulfated product may be formed from an unsulfated precursor molecule, when this is produced in vivo.

Furthermore, the data surprisingly show that employing the tyrosine ammonia-lyase from *Flavobacterium johnsoniae* (FjTAL; SEQ ID NO: 14) results in a higher supply in unsulfated precursor molecule (here: p-coumaric acid), which in turn leads to a higher yield of sulfated product (here: zosteric acid) compare to other tyrosine ammonia-lyases.

Example 6 – Production of sulfated products in other hosts

We have shown that zosteric acid can be produced in vivo in *Escherichia coli* by expression of an aryl sulfotransferase. To show that the reaction is possible in other microorganisms, we here show that the yeast *Saccharomyces cerevisiae* can also be used as a host for the production.

The gene encoding aryl sulfotransferase SULT1A (Example 1) was cloned after a TEF1 promoter into an episomal plasmid with a 2-micron origin of replication as follows. The gene was amplified by PCR using primers CBJP633 and CBJP634. Alternatively, the gene was codon-optimized for *E. coli* and synthesized by GeneArt and amplified by primers CBJP635 and CBJP636. The TEF1 promoter (Jensen et al., 2014, *FEMS Yeast Res* 14: 238-248) was amplified by PCR using the primers PTEF1_fw and PTEF1_rv. Plasmid pCfB132 (Jensen et al., supra) was digested by restriction enzymes AsiSI and Nt.BsmI. The three fragments – plasmid, TEF1 promotor and SULT1A1-encoding gene – were assembled using a uracil-excission cloning procedure, resulting in plasmids pCBJ283 and pCBJ284 (Figures 9 and 10, respectively, which was subsequently transformed into the *Saccharomyces cerevisiae* strain CEN.PK102-5B selecting for growth on synthetic dropout media plates lacking uracil. A control strain was also made by transformation of pCfB132 into CEN.PK102-5B.

Primers:

Oligonucleotide	Gene/promoter	Direction	Sequence
CBJP633	SULT1A1 rat	Forward	AGTGCAGGUAAAACAATGgagttctcccgtcca
CBJP634	SULT1A1 rat	Reverse	CGTGCGAUTCAtagttcacaacgaaacttg
CBJP635	SULT1A1 rat (E. coli)	Forward	ATCTGTCAUAAAACAATGgaattttcacgtccgc
CBJP636	SULT1A1 rat (E. coli)	Reverse	CACGCGAUTCAcagttcacaacgaaatttgaa
PTEF1_fw	PTEF1	Forward	Cacgcgaugcacacaccatagcttc
PTEF1_rv	PTEF1	Reverse	Cgtgcgauggaagtaccttcaaaga

The strains were grown in modified Delft medium (Jensen et al., supra) with 20 mg/mL histidine and 60 mg/mL leucine and 10 mM *p*-coumaric acid overnight at 30°C with aeration. The supernatant was then isolated and examined by HPLC as described in Example 1. The table below shows that zosteric acid (ZA) was produced by the strain expressing SULT1A1 and not the control strain lacking a sulfotransferase.

Sulfotransferase	**µM ZA** (averages and standard deviations of replicate experiments)
None	0 ± 0
SULT1A1 rat (native)	37.8 ± 5.7
SULT1A1 rat (codon optimized for E. coli)	46.2 ± 3.5

It is evident that zosteric acid is formed only when a sulfotransferase is expressed in yeast, and that the gene encoding this may be natural or encoded by a synthetic gene with a specific codon-optimization. Conclusively, the sulfation reactions shown to be catalyzed by sulfotransferases in *E. coli* are also catalyzed when the sulfotransferases are expressed in other organisms, as demonstrated here for the yeast *S. cerevisiae*. The efficacy of production may be affected by means such as the codon-usage of the genes encoding the sulfotransferase. Thus yeast expressing sulfotransferases may be able to detoxify aromatic compounds such as *p*-coumaric acid, and form sulfated products such as zosteric acid.

Example 7 – A range of compounds are substrates for sulfation in vivo

Here we show that the expression of an aryl sulfotransferase may be able to convert several substrates. Some of these are inhibitors that can be found in biomass hydrolyzate used as a substrate for cell growth and production in biotechnology. The compounds also include some that are of biotechnological interest as products of a cell culture or be some whose sulfate ester is of economic interest.

Different sulfotransferases were examined for their substrate specificities against three substrates. We tested the sulfotransferases mentioned in example 1, as well as additional ones. The genes encoding these were cloned as described in example 1 using the primers shown in the table below from cDNA libraries of the respective organisms, except for the SULT1A1 from rat (*Rattus norvegicus*) codon-optimized for *E. coli* (described above). The resulting vectors were transformed into BL21(DE3)pLysS.

Primers:

Oligonucleotide	**Gene**	**Direction**	**Sequence**
CBJP517	SULT1C1 *Gallus gallus domesticus*	Forward	TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCatggccctggataaaatgg
CBJP518	SULT1C1 *Gallus gallus domesticus*	Reverse	TAAGCATTATGCGGCCGCAAGCTTGtcacaattccatgcgaaaaactag
CBJP533	SULT1A1 *Rattus norvegicus* (Codon-optimized for E. coli)	Forward	TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCatggaattttcacgtcc
CBJP534	SULT1A1 *Rattus norvegicus* (Codon-optimized for E. coli)	Reverse	TAAGCATTATGCGGCCGCAAGCTTGttacagttcacaacgaaatttg

The resulting strains were grown in M9 medium containing either 100 µM pHCA, 95 µM resveratrol or 87 µM kaempferol. The cultures were grown overnight at 37°C, 300 rpm. The following day the supernatants were isolated and examined by HPLC as described in example 1. BL21(DE3)pLysS were used as a control strain and did not convert the substrates.

Enzyme	pHCA	resveratrol	kaempferol
	100 µM	95 µM	87 µM
SULT1A1 *Rattus norvegicus*	93%	93%	95%
SULT1C1 *Gallus gallus domesticus*	26%	100%	80%
SULT1A1 *Rattus norvegicus* (Codon-optimized for *E. coli*)	73%	58%	38%
SULT1A1 human	39%	36%	97%
SULT1A1 *Equus caballus*	21%	100%	96%
SULT1E1 *Gallus gallus domesticus*	17%	100%	47%
SULT1A1 *Canis lupus familiaris*	34%	61%	60%
SULT1A1 *Sus scrofa domesticus*	8%	88%	45%

The table shows the percent conversion of the various substrates by cells expressing the different sulfotransferases. The results show that several sulfotransferases, and especially the aryl sulfotransferase from rat (*Rattus norvegicus*), may be employed in the sulfation of phenolic compounds.

To further test the range of substrates that can be sulfated, we used strains carrying plasmids expressing SULT1A1 from rat (*Rattus norvegicus*) and SULT1E1 from chicken (*Gallus gallus domesticus*) (Example 1) cloned into the expression vector pETDuet-1, and cysDNCQ from *E. coli* cloned into expression vector pRSFDuet-1 (Example 2). The plasmids were introduced into the *E. coli* expression strain BL21(DE3)pLysS as described previously, selecting for transformants with appropriate antibiotics, namely 34 µg mL^{-1} chloramphenicol for pLysS, 100 µg mL^{-1} ampicillin for pETDuet-1-based vectors, and 100 µg mL^{-1} kanamycin for pRSFDuet-1-based vectors. The table below shows the combination of over-expressed genes on plasmids. A control strain without a sulfotransferase gene or cysDNCQ operon was also examined.

E. coli strains	Sulfotransferase	Cys genes
Control strain	-	-
SULT1A1 rat	SULT1A1 rat	-
SULT1E1 chicken	SULT1E1 chicken	-
SULT1A1 rat + CysDNCQ	SULT1A1 rat	CysDNCQ

The strains were precultured in 2xYT medium with appropriate antibiotics. 10 µL of these precultures were used to inoculate M9 media with 1 mM IPTG and none or a single substrate for sulfation. After overnight growth at 37°C, 300 rpm the supernatants were withdrawn and examined by HPLC as described in Example 1. The compounds were detected by UV absorbance. The table below shows the percent reduction in concentration in the strains expressing sulfotransferases alone or in combination with cysDNCQ genes when compared to the control strain.

Compound	Start concentration in µM	SULT1A1	SULT1E1	SULT1A1 + CysDNCQ
Ferulic acid	110	72%	67%	100%
Quercetin	85	75%	74%	81%
4-hydroxybenzoic acid	287	5%	4%	6%
4-acetamidophenol	114	24%	10%	30%
3-Hydroxy-4-methoxycinnamic acid	132	51%	24%	62%
4-Hydroxyphenylpyruvic acid	255	47%	100%	64%
3-(4-Hydroxyphenyl)propionic acid	241	3%	1%	7%
Vanillic acid	173	33%	0%	39%
Luteolin	61	27%	0%	37%
Apigenin	77	41%	98%	99%
fisetin	81	98%	98%	100%

Conclusively, a wide range of phenolic compounds are substrates for sulfotransferases. In the shown examples, the conversion is enhanced by the overexpression of cysDNCQ genes. Some of these compounds and their sulfate esters are of interest in biotechnology. Also, some of these compounds are inhibitors of cell growth and function, and thus conversion by sulfation is of interest for use in biological systems.

Embodiments of the invention

1. A process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

2. The process according to item 1, further comprising:

(ii) culturing the first recombinant host cell under suitable conditions for the production of the corresponding sulfated phenolic compound; and

(iii) optionally, recovering said sulfated phenolic compound.

3. The process according to item 1 or 2, wherein the heterologous polypeptide having an aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme.

4. The process according to any one of items 1-3, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

5. The process according to any one of items 1-4, wherein the heterologous polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

6. The process according to any one of items 1-5, wherein the first recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide.

7. The process according to item 6, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide.

8. The process according to item 6 or 7, wherein the exogenous nucleic acid molecule is a vector.

9. The process according to item 6 or 7, wherein the exogenous nucleic acid molecule is stabily integrated into the genome of said first recombinant host cell.

10. The process according to any one of items 1-7, wherein the first recombinant host cell has been further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

11. The process according to item 10, wherein the ATP sulfurylase is encoded by the genes cysD and cysN.

12. The process according to any one of items 1-11, wherein said first recombinant host cell has been further modified to have an increased poretin expression of an APS kinase compared to an identical host cell that does not carry said modification.

13. The process according to item 12, wherein the said APS kinase is encoded by the gene cysC.

14. The process according to any one of items 1-13, wherein said first recombinant host cell has been further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

15. The process according to item 14, wherein said PAP phosphatase is encoded by the gene cycQ.

16. The process according to any one of items 10-15, wherein the increase in protein expression is achieved by increasing the number of copies of the encoding gene or genes.

17. The process according to item 16, wherein the increase in the number of copies of the gene or genes is achieved by using one or more vectors comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

18. The process according to any one of items 10-15, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

19. The process according to any one of items 10-18, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes.

20. The process according to any one of items 1-19, wherein said first recombinant host cell further comprises a heterologous polypeptide having a tyrosine ammonia lyase activity.

21. The process according to any one of items 1-20, wherein in step (i'), (i'') or (i''') the medium is further contacted with a second recombinant host cell comprising a heterologous polypeptide having a tyrosine ammonia lyase activity.

22. The process according to item 20 or 21, wherein the heterologous polypeptide having a tyrosine ammonia lyase activity is selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

23. The process according to any one of items 20 to 22, wherein the first and/or second recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide having a tyrosine ammonia lyase activity.

24. The process according to item 23, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide.

25. The process according to item 23 or 24, wherein the exogenous nucleic acid molecule is a vector.

26. The process according to item 23 or 24, wherein the exogenous nucleic acid is stably integrated into the genome of the first and/or second recombinant host cell.

27. The process according to any one of items 1 to 26, wherein the first recombinant host cell and the second recombinant host cell are independently selected from the group consisting of bacteria, yeasts, fungi, algae and plant cells.

28. The process according to any one of items 1 to 27, wherein the first recombinant host cell is a bacterium.

29. The process according to item 28, wherein the bacterium is a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus,* or *Yersinia*.

30. The process according to item 28, wherein the bacterium is a bacterium of the genus *Bacillus*.

31. The process according to item 30, wherein the bacterium is *Bacillus subtilis*.

32. The process according to item 28, wherein the bacterium is a bacterium of the genus *Lactococcus*.

33. The process according to item 32, wherein the bacterium is *Lactococcus lactis*.

34. The process according to item 28, wherein the bacterium is a bacterium of the genus *Pseudomonas*.

35. The process according to item 34, wherein the bacterium is *Pseudomonas putida*.

36. The process according to item 28, wherein the bacterium is a bacterium of the genus *Corynebacterium*.

37. The process according to item 36, wherein the bacterium is *Corynebacterium glutamicum*.

38. The process according to item 28, wherein the bacterium is a bacterium of the genus *Escherichia*.

39. The process according to item 38, wherein the bacterium is *Escherichia coli*.

40. The process according to any one of item 1-27, wherein the first recombinant host cell is a yeast.

41. The process according to item 40, wherein the yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces,*

Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula, or *Trichosporon*.

42. The process according to item 40, wherein the yeast is a yeast of the genus *Saccharomyces* or *Pichia*.

43. The process according to item 40, wherein the yeast is selected from the group consisting of *Saccharomyces cerevisiae, Pichia pastoris*, and P*ichia kudriavzevii*.

44. The process according to item 43, wherein the yeast is *Saccharomyces cerevisiae*.

45. The process according to item 43, wherein the yeast is *Pichia pastoris*.

46. The process according to any one of items 1-27, wherein the first recombinant host cell is a fungus.

47. The process according to item 46, wherein the fungus is a fungus of the genus *Aspergillus*.

48. The process according to item 47, wherein the fungus is *Aspergillus Oryzae* or *Aspergillus niger*.

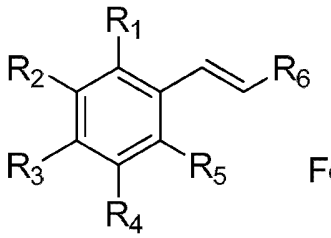
49. The process according to any one of items 1-27, wherein the first recombinant host cell is an algae cell.

50. The process according to item 49, wherein the algae cells is an algae cell of the genus *Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

51. The process according to any one of items 1-27, wherein the first recombinant host cell is a plant cell.

52. The process according to item 51, wherein the plant cell is selected from the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

53. The process according to any one of items 1-52, wherein the phenolic compound is represented by the general formula (I):



Formula (I)

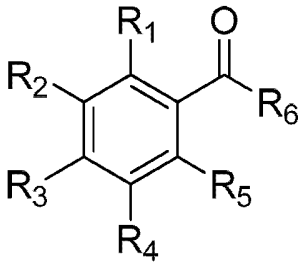
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

54. A process according to any one of the items 1-52, wherein the phenolic compound is represented by the general formula (II):



Formula (II)

;

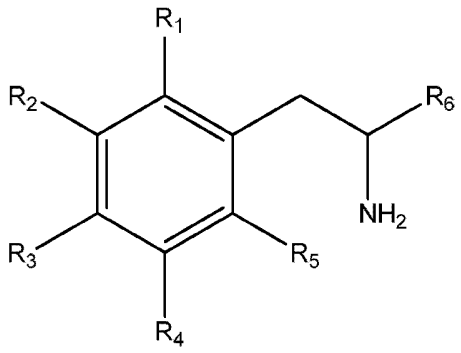
wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are

independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

55. The process according to any one of items 1-53, wherein the precursor of a phenolic compound in step (i''') is a compound of the general Formula (p-I):



Formula (p-I);

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

56. The process according to any one of items 53-55, wherein R_6 is -$COOR_7$, wherein R_7 is selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl.

57. The process according to item 56, wherein R_7 is hydrogen.

58. The process according to any one of items 53-57, wherein R_3 is hydroxyl (-OH).

59. The process according to any one of items 53-58, wherein each of R_1, R_2, R_4 and R_5 is hydrogen.

60. The process according to any one of items 53-58, wherein R_4 is hydroxyl (-OH).

61. The process according to item 60, wherein each of R_1, R_2, and R_5 is hydrogen.

62. The process according to any one of items 53-55, wherein each of R_1, R_3 and R_5 is hydrogen, each of R_2 and R_4 is hydroxyl (-OH), and R_6 is p-hydroxyphenyl.

63. A recombinant host cell comprising a first heterologous polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

64. The recombinant host cell according to item 63, wherein the heterologous polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

65. The recombinant host cell according to item 63 or 64, wherein the polypeptide according to b) or c) has aryl sulfotransferase activity.

66. The recombinant host cells according to any one of items 63-65, the host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said first heterologous polypeptide.

67. The recombinant host cell according to item 66, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said first heterologous polypeptide.

68. The recombinant host cell according to item 67, wherein the exogenous nucleic acid molecule further comprises at least one regulatory element selected from a 5' untranslated region (5'UTR) and 3' untranslated region (3' UTR).

69. The recombinant host cell according to any one of items 66-68, wherein the exogenous nucleic acid is a vector.

70. The recombinant host cell according to any one of items 66-68, wherein the excogenous nucleic acid is stably integrated into the genome of the host cell.

71. The recombinant host cell according to any one of items 63-70, wherein the recombinant host cell has further been modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

72. The recombinant host cell according to item 71, wherein said ATP sulfurylase is encoded by the genes cysD and cysN.

73. The recombinant host cell according to any one of items 63-72, wherein the recombinant host cell has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

74. The recombinant host cell according to item 73, wherein said APS kinase is encoded by the gene cysC.

75. The recombinant host cell according to any one of items 63-74, wherein the recombinant host cell has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

76. The recombinant host cell according to item 75, wherein said PAP phosphatase is encoded by the gene cycQ.

77. The recombinant host cell according to any one of items 63-76, wherein the increase in gene expression has been achieved by an increased number of copies of the gene or genes.

78. The precombinant host cell according to item 77, wherein the increase in the number of copies of the gene or genes is achieved by using one or more vectors comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

79. The recombinant host cell according to any one of item 63-76, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

80. The recombinaint host cell according to any one of items 63-76, wherein the increase in gene expression has been achieved by increasing the strength of the promoter(s) operably linked to the gene or genes.

81. The recombinant host cell according to any one of items 63-80, further comprising a second heterologous polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least

about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

82. The recombinant host cell according to item 81, wherein the heterologous polypeptide according to e) or f) has tyrosine ammonia lyase activity.

83. The recombinant host cell according to item 81 or 82, wherein the recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said second heterologous polypeptide.

84. The recombinant host cell according to item 83, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said second heterologous polypeptide.

85. The recombinant host cell according to item 83 or 84, wherein the exogenous nucleic acid molecule is a vector.

86. The recombinant host cell according to item 83 or 84, wherein the exogenous nucleic acid is stably integrated into the genome of the recombinant host cell.

87. The recombinant host cell according to any one of items 63-86, wherein the recombinant host cell is selected from the group consisting of bacteria, yeasts, fungi, algae and plant cells.

88. The recombinant host cell according to any one of items 63-86, wherein the recombinant host cell is a bacterium.

89. The recombinant host cell according to item 88, wherein the bacterium is a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter,*

Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus, or *Yersinia*.

90. The recombinant host cell according to item 88, wherein the bacterium is a bacterium of the genus *Bacillus*.

91. The recombinant host cell according to item 90, wherein the bacterium is *Bacillus subtilis*.

92. The recombinant host cell according to item 88, wherein the bacterium is a bacterium of the genus *Lactococcus*.

93. The recombinant host cell according to item 92, wherein the bacterium is *Lactococcus lactis*.

94. The recombinant host cell according to item 88, wherein the bacterium is a bacterium of the genus *Pseudomonas*.

95. The recombinant host cell according to item 94, wherein the bacterium is *Pseudomonas putida*.

96. The recombinant host cell according to item 88, wherein the bacterium is a bacterium of the genus *Corynebacterium*.

97. The recombinant host cell according to item 96, wherein the bacterium is *Corynebacterium glutamicum*.

98. The recombinant host cell according to item 88, wherein the bacterium is a bacterium of the genus *Escherichia*.

99. The recombinant host cell according to item 98, wherein the bacterium is *Escherichia coli*.

100. The recombinant host cell according to any one of items 63-86, wherein the recombinant host cell is a yeast.

101. The recombinant host cell according to item 100, wherein the yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen,*

Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula, or *Trichosporon*.

102. The recombinant host cell according to item 100, wherein the yeast is a yeast of the genus *Saccharomyces* or *Pichia*.

103. The recombinant host cell according to item 100, wherein the yeast is selected from the group consisting of *Saccharomyces cerevisiae, Pichia pastoris*, and P*ichia kudriavzevii*.

104. The recombinant host cell according to item 103, wherein the yeast is *Saccharomyces cerevisiae*.

105. The recombinant host cell according to item 103, wherein the yeast is *Pichia pastoris*.

106. The recombinant host cell accoding to any one of items 63-86, wherein the recombinant host cell is a fungus.

107. The recombinant host cell according to item 106, wherein the fungus is a fungus of the genus *Aspergillus*.

108. The recombinant host cell according to item 107, wherein the fungus is *Aspergillus Oryzae* or *Aspergillus niger*.

109. The recombinant host cell according to any one of items 63-86, wherein the recombinant host cell is an algae cell.

110. The recombinant host cell according to item 109, wherein the algae cells is an algae cell of the genus *Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

111. The recombinant host cell according to any one of items 63-86, wherein the recombinant host cell is a plant cell.

112. The recombinant host cell according to item 111, wherein the plant cell is selected from the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

113. The recombinant host cell according to any one of items 63-112, which is employed as first recombinant host cell in the process according to any one of items 1-62.

114. Use of a polypeptide in the sulfation of a phenolic compound, said polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

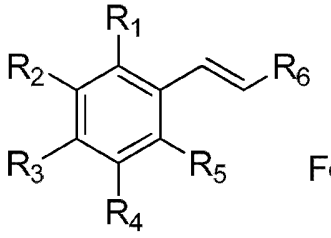
115. The use according to item 114, wherein the polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

116. The use according to item 114 or 115, wherein the phenolic compound is of the general formula (I):



Formula (I)

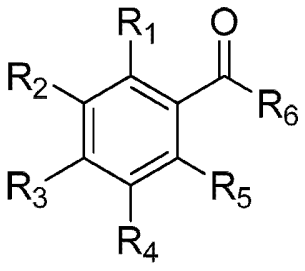
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

117. The use according to item 114 or 115, wherein the phenolic compound is of the general formula (II)



Formula (II)

;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are

independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

118. The use according to item 114 or 115, wherein the phenolic compound is p-coumaric acid.

119. Process for the production of a sulfated phenolic compound, such as zosteric acid, the method comprises sulfating a phenolic compound, such as p-coumaric acid, using a polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

120. The process according to item 119, wherein the polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least

about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

121. The process according to item 119 or 120, wherein the phenolic compound is of the general formula (I) or (II) as defined herein.

122. A composition comprising a first recombinant host cell comprising a heterologous polypeptide having arylsulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted; and

a second recombinant host cell comprising a heterologous polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence

identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

123. A composition comprising a first polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted; and

a second polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14);

e) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14); or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 (e.g., SEQ ID NO: 14), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

5

Claims

1. A process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity.

2. The process according to claim 1, further comprising:

(ii) culturing the first recombinant host cell under suitable conditions for the production of the corresponding sulfated phenolic compound; and

(iii) optionally, recovering said sulfated phenolic compound.

3. The process according to claim 1 or 2, wherein the heterologous polypeptide having an aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme.

4. The process according to any one of claims 1-3, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

5. The process according to any one of claims 1-4, wherein the heterologous polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

6. The process according to any one of claims 1-5, wherein the first recombinant host cell has been further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

7. The process according to claim 6, wherein the ATP sulfurylase is encoded by the genes cysD and cysN.

8. The process according to any one of claims 1-7, wherein said first recombinant host cell has been further modified to have an increased poretin expression of an APS kinase compared to an identical host cell that does not carry said modification.

9. The process according to claim 8, wherein the said APS kinase is encoded by the gene cysC.

10. The process according to any one of claims 1-9, wherein said first recombinant host cell has been further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

11. The process according to claim 10, wherein said PAP phosphatase is encoded by the gene cycQ.

12. The process according to any one of claims 1-11, wherein said first recombinant host cell further comprises a heterologous polypeptide having a tyrosine ammonia lyase activity.

13. The process according to any one of claims 1-12, wherein in step (i'), (i'') or (i''') the medium is further contacted with a second recombinant host cell comprising a heterologous polypeptide having a tyrosine ammonia lyase activity.

14. The process according to claim 12 or 13, wherein the heterologous polypeptide having a tyrosine ammonia lyase activity is selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

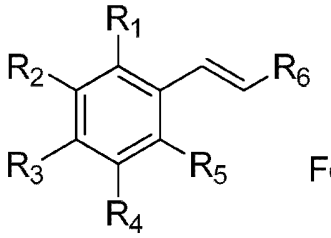
15. The process according to claim 12 or 13, wherein the heterologous polypeptide having a tyrosine ammonia lyase activity is selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

16. The process according to any one of claims 1-15, wherein the phenolic compound is represented by the general formula (I):



Formula (I)

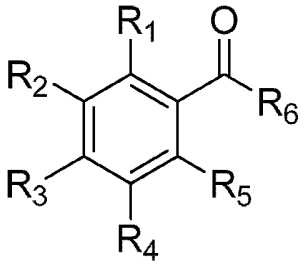
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

17. A process according to any one of claims 1-15, wherein the phenolic compound is represented by the general formula (II):



Formula (II)

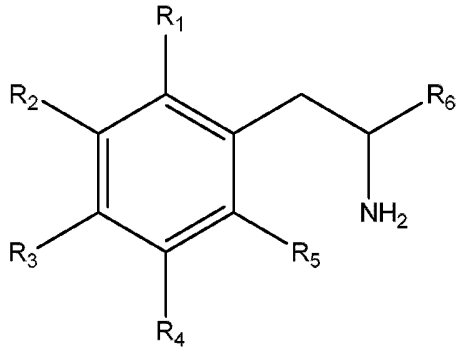
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

18. The process according to any one of claims 1-15, wherein the precursor of a phenolic compound in step (i''') is a compound of the general Formula (p-I):



Formula (p-I);

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

19. The process according to any one of claims 16-18, wherein R_6 is $-COOR_7$ and R_7 is hydrogen.

20. The process according to any one of claims 16-19, wherein R_3 is hydroxyl (-OH).

21. The process according to any one of claims 16-20, wherein each of R_1, R_2, R_4 and R_5 is hydrogen, R_3 is hydroxyl (-OH) and R_6 is -COOH .

22. A recombinant host cell comprising a first heterologous polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

23. The recombinant host cell according to claim 22, wherein the heterologous polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

24. The recombinant host cell according to claim 22 or 23, wherein the recombinant host cell has further been modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

25. The recombinant host cell according to claim 24, wherein said ATP sulfurylase is encoded by the genes cysD and cysN.

26. The recombinant host cell according to any one of claims 22 to 25, wherein the recombinant host cell has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

27. The recombinant host cell according to claim 26, wherein said APS kinase is encoded by the gene cysC.

28. The recombinant host cell according to any one of claims 22 to 27, wherein the recombinant host cell has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

29. The recombinant host cell according to claim 28, wherein said PAP phosphatase is encoded by the gene cycQ.

30. The recombinant host cell according to any one of claims 22 to 29, further comprising a second heterologous polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

31. The recombinant host cell according to any one of claims 22 to 29, further comprising a second heterologous polypeptide having tyrosine ammonia lyase activity, wherein the heterologous polypeptide is selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

32. The recombinant host cell according to any one of claims 22-31, which is employed as first recombinant host cell in the process according to any one of claims 1-21.

33. Process for the production of a sulfated phenolic compound, such as zosteric acid, the method comprises sulfating a phenolic compound, such as p-coumaric acid, using a polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

34. The process according to claim 33, wherein the polypeptide is selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity.

35. A composition comprising a first recombinant host cell comprising a heterologous polypeptide having arylsulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity; and

a second recombinant host cell comprising a heterologous polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

36. The composition according to claim 35, comprising a first recombinant host cell comprising a heterologous polypeptide having arylsulfotransferase activity selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity; and

a second recombinant host cell comprising a heterologous polypeptide having tyrosine ammonia lyase activity selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

37. A composition comprising a first polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity; and

a second polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

38. The composition according to claim 37, comprising a first polypeptide having aryl sulfotransferase activity selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1;

b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, wherein the polypeptide has aryl sulfotransferase activity; or

c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has aryl sulfotransferase activity; and

a second polypeptide having tyrosine ammonia lyase activity selected from the group consisting of:

d) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

e) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 14, wherein the polypeptide has tyrosine ammonia lyase activity; or

f) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, wherein the polypeptide has tyrosine ammonia lyase activity.

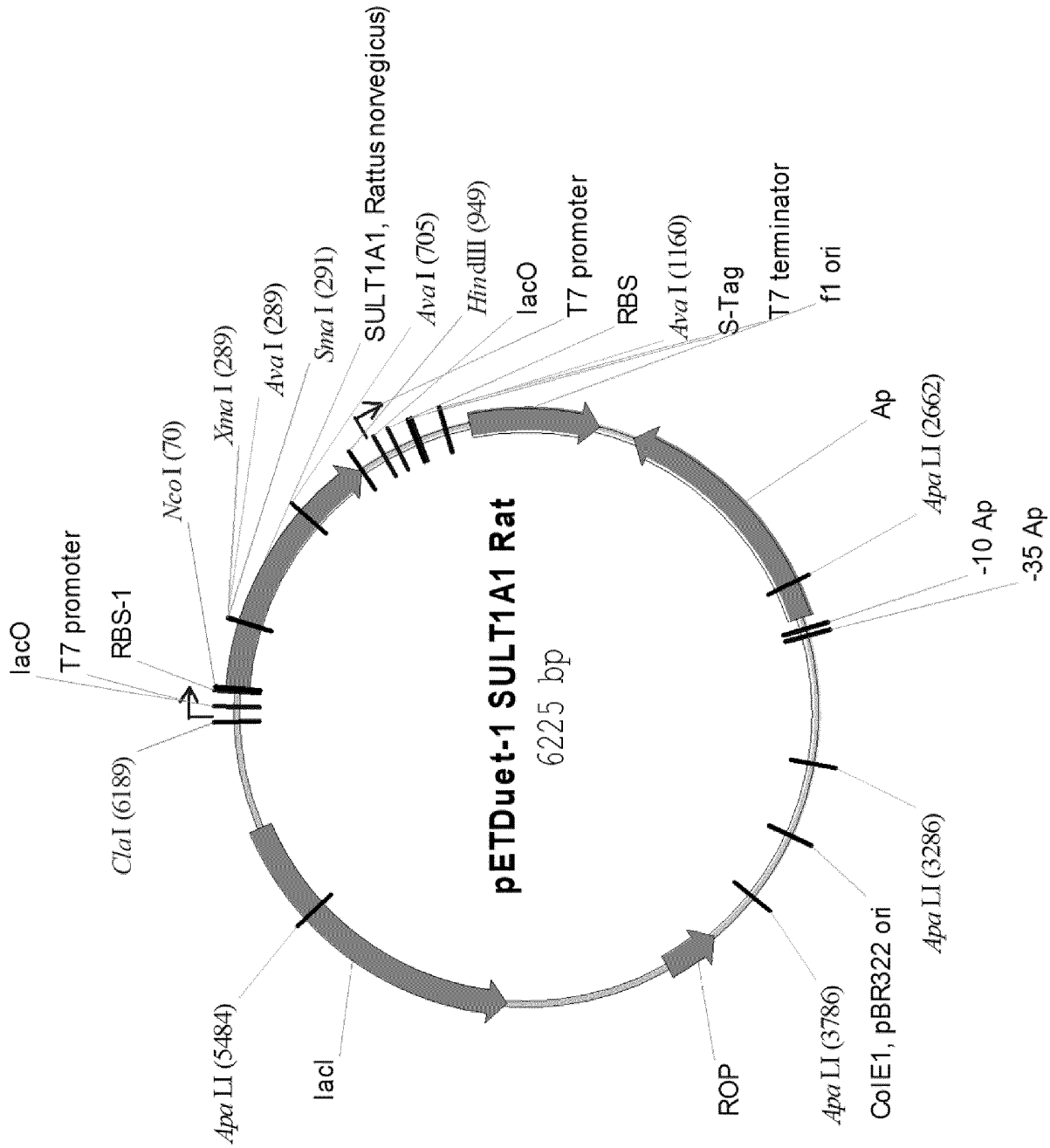

pETDuet-1 SULT1A1 Rat
6225 bp
lacO
T7 promoter
RBS-1
ClaI (6189)
NcoI (70)
XmaI (289)
AvaI (289)
SmaI (291)
SULT1A1, Rattus norvegicus)
AvaI (705)
HindIII (949)
lacO
T7 promoter
RBS
AvaI (1160)
S-Tag
T7 terminator
f1 ori
Ap
ApaLI (2662)
-10 Ap
-35 Ap
ApaLI (3286)
ColE1, pBR322 ori
ApaLI (3786)
ROP
lacI
ApaLI (5484)

Figure 1

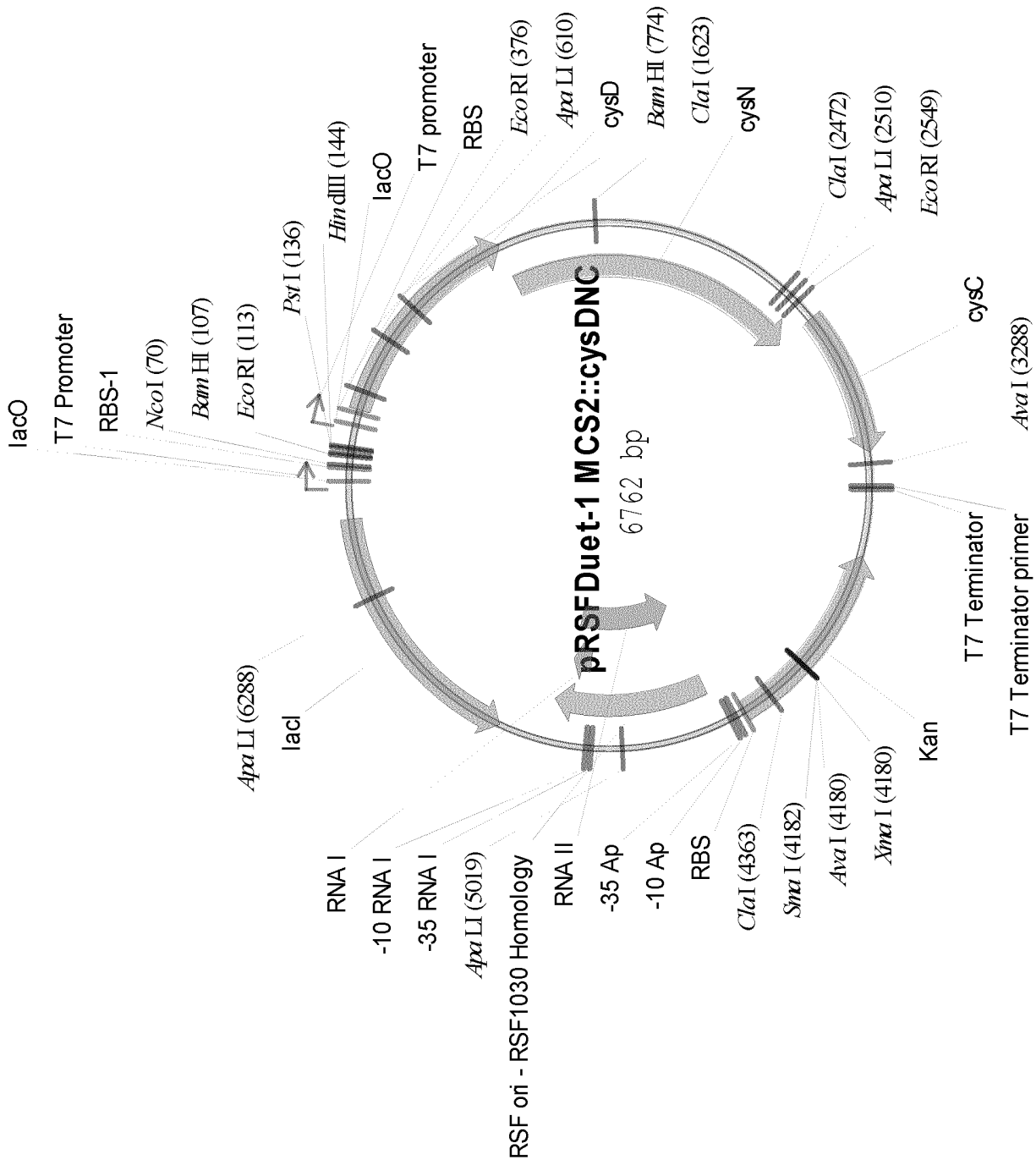

pRSFDuet-1 MCS2::cysDNC
6762 bp
lacO
T7 Promoter
RBS-1
NcoI (70)
BamHI (107)
EcoRI (113)
PstI (136)
HindIII (144)
lacO
T7 promoter
RBS
EcoRI (376)
ApaLI (610)
cysD
BamHI (774)
ClaI (1623)
cysN
ClaI (2472)
ApaLI (2510)
EcoRI (2549)
cysC
AvaI (3288)
T7 Terminator
T7 Terminator primer
Kan
XmaI (4180)
AvaI (4180)
SmaI (4182)
ClaI (4363)
RBS
-10 Ap
-35 Ap
RNA II
RSF ori - RSF1030 Homology
ApaLI (5019)
-35 RNA I
-10 RNA I
RNA I
lacI
ApaLI (6288)

Figure 2

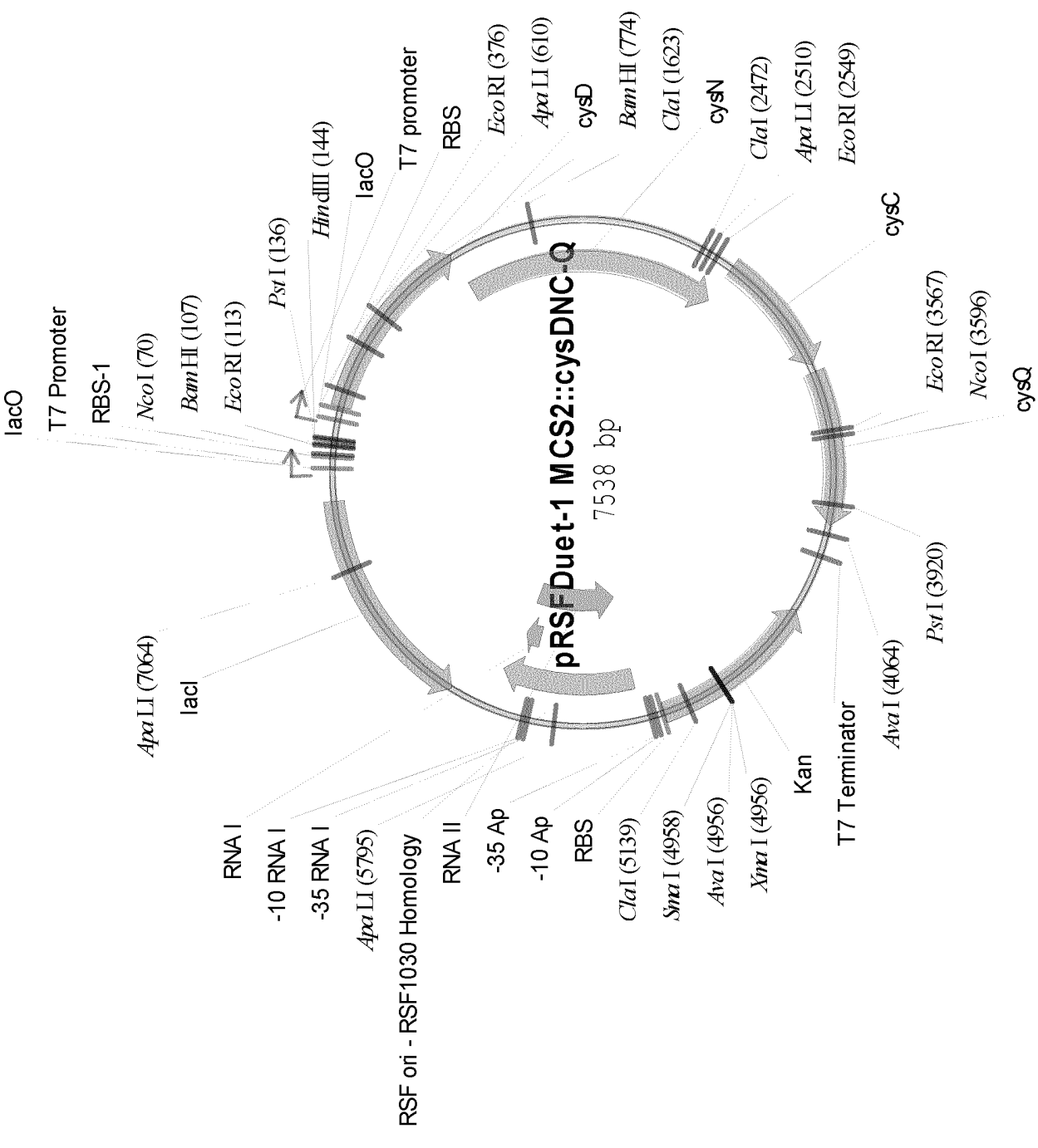

pRSFDuet-1 MCS2::cysDNC-Q
7538 bp
lacO
T7 Promoter
RBS-1
NcoI (70)
BamHI (107)
EcoRI (113)
PstI (136)
HindIII (144)
lacO
T7 promoter
RBS
EcoRI (376)
ApaLI (610)
cysD
BamHI (774)
ClaI (1623)
cysN
ClaI (2472)
ApaLI (2510)
EcoRI (2549)
cysC
EcoRI (3567)
NcoI (3596)
cysQ
PstI (3920)
AvaI (4064)
T7 Teminator
Kan
XmaI (4956)
AvaI (4956)
SmaI (4958)
ClaI (5139)
RBS
-10 Ap
-35 Ap
RNA II
RSF ori - RSF1030 Homology
ApaLI (5795)
-35 RNA I
-10 RNA I
RNA I
lacI
ApaLI (7064)

Figure 3

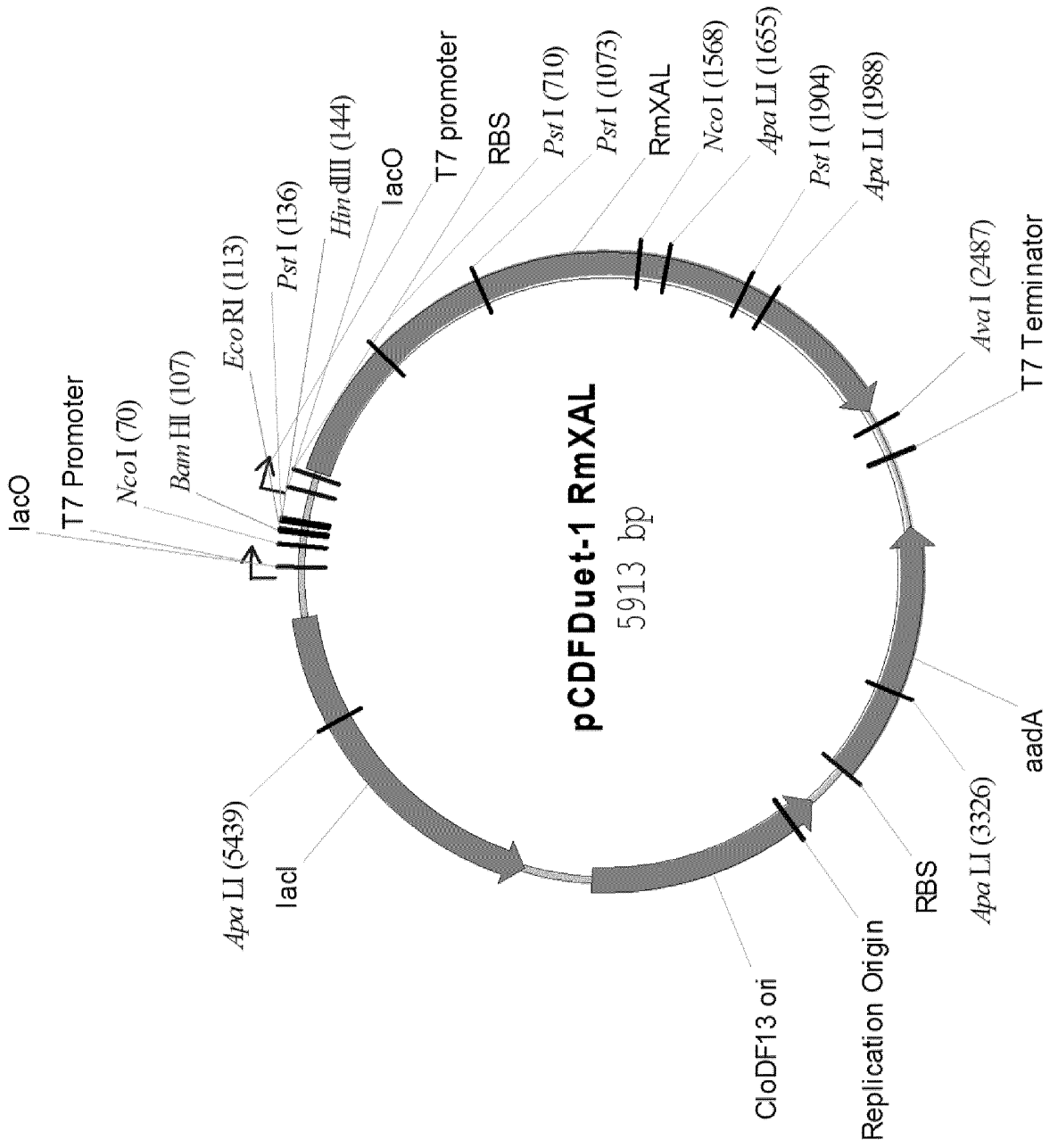

pCDFDuet-1 RmXAL
5913 bp
lacO
T7 Promoter
NcoI (70)
BamHI (107)
EcoRI (113)
PstI (136)
HindIII (144)
lacO
T7 promoter
RBS
PstI (710)
PstI (1073)
RmXAL
NcoI (1568)
ApaLI (1655)
PstI (1904)
ApaLI (1988)
AvaI (2487)
T7 Terminator
aadA
ApaLI (3326)
RBS
Replication Origin
CloDF13 ori
lacI
ApaLI (5439)

Figure 4

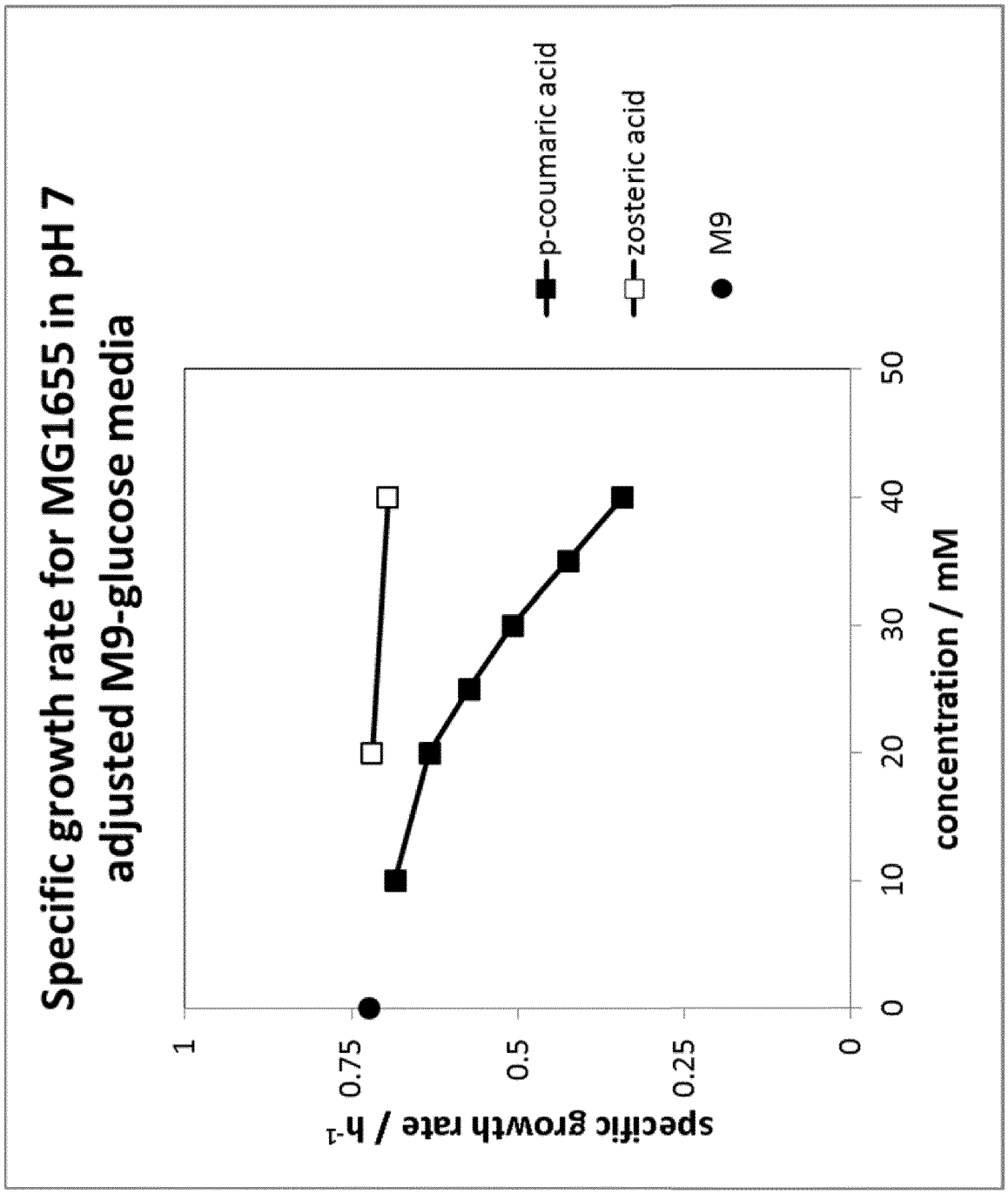

Specific growth rate for MG1655 in pH 7
adjusted M9-glucose media
p-coumaric acid
zosteric acid
M9
specific growth rate / h-1
concentration / mM
0
0.25
0.5
0.75
1
0
10
20
30
40
50

Figure 5

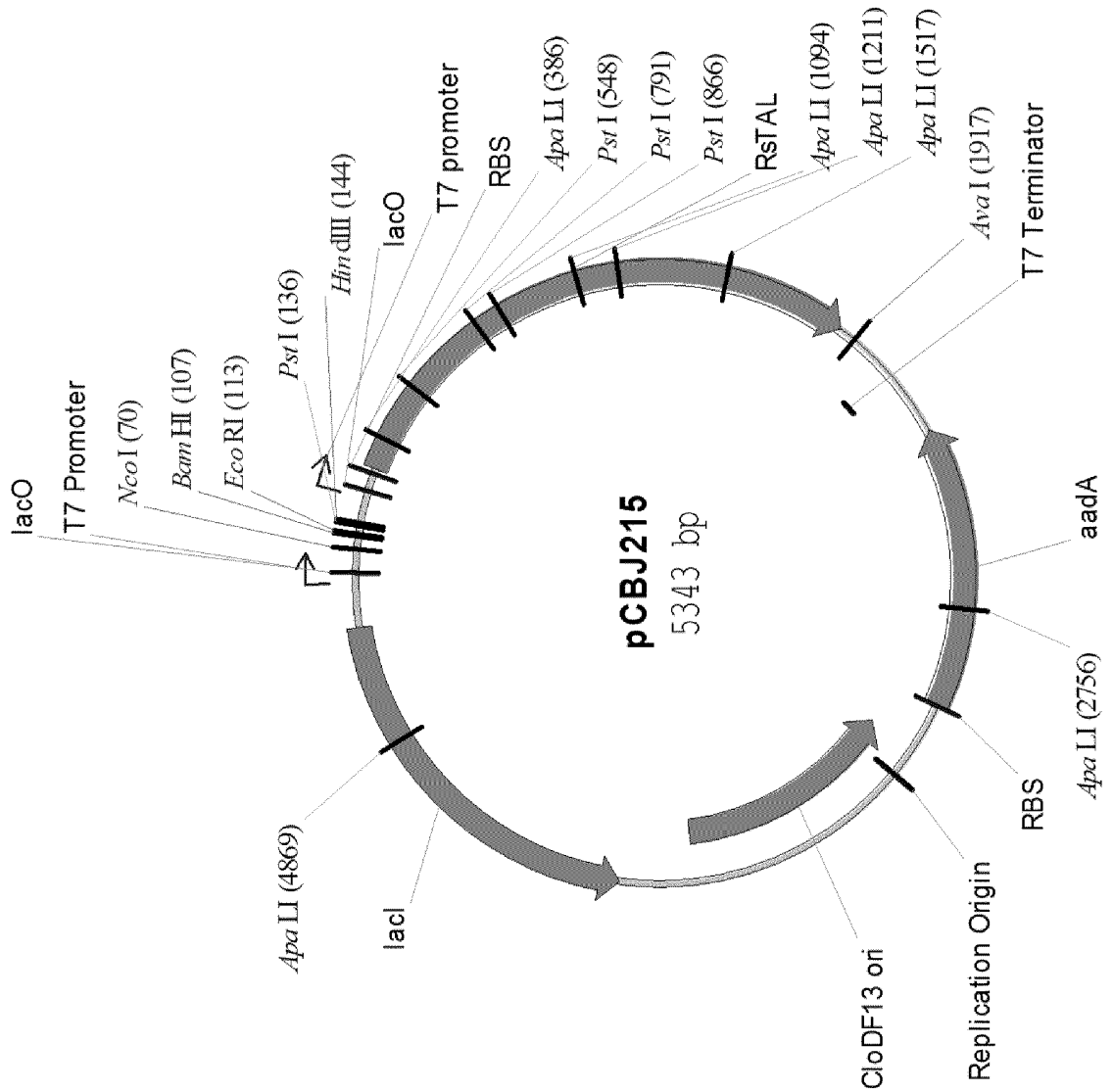

pCBJ215
5343 bp
lacO
T7 Promoter
NcoI (70)
BamHI (107)
EcoRI (113)
PstI (136)
HindIII (144)
lacO
T7 promoter
RBS
ApaLI (386)
PstI (548)
PstI (791)
PstI (866)
RsTAL
ApaLI (1094)
ApaLI (1211)
ApaLI (1517)
AvaI (1917)
T7 Terminator
aadA
ApaLI (2756)
RBS
Replication Origin
CloDF13 ori
lacI
ApaLI (4869)

Figure 6

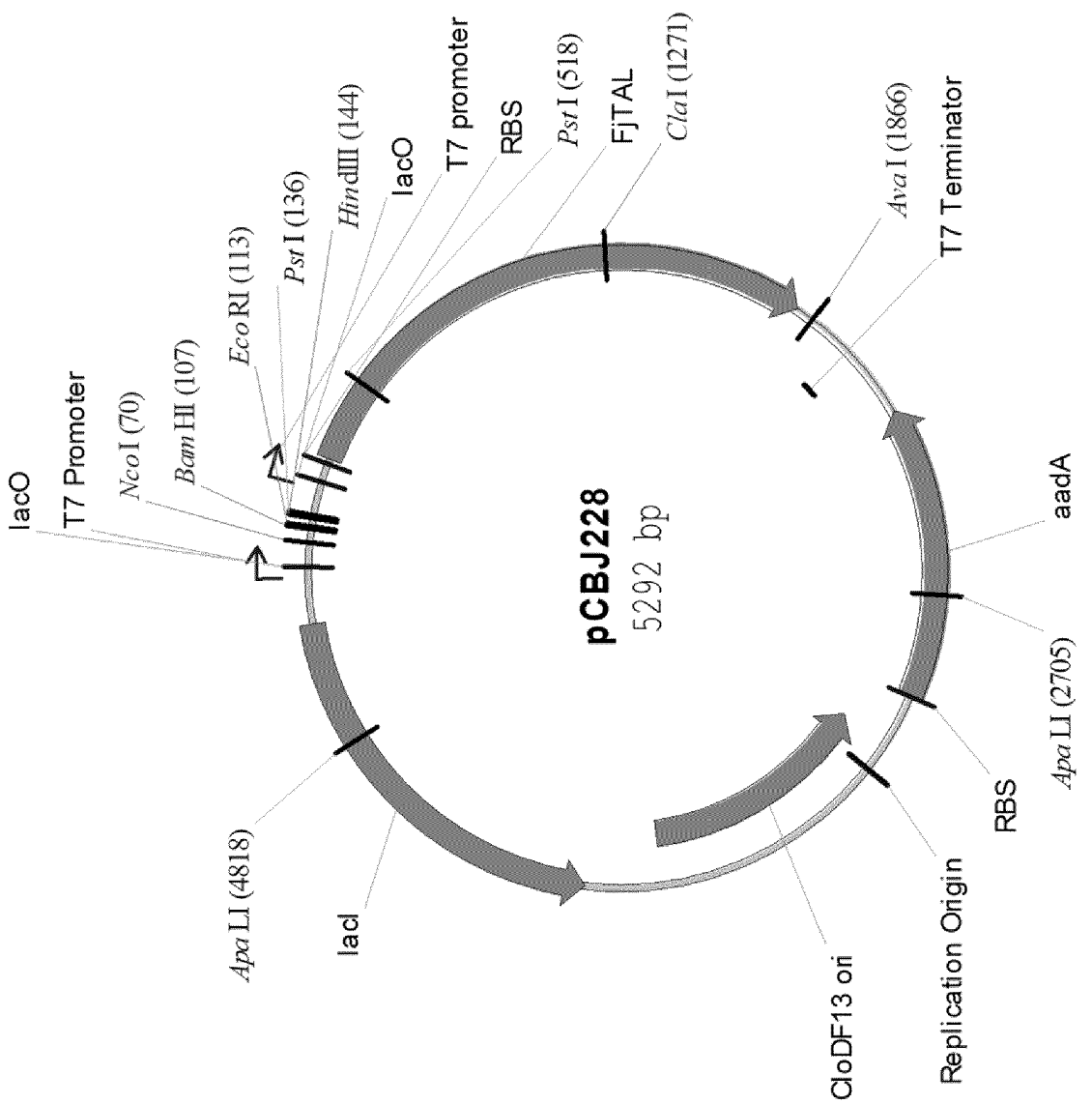

pCBJ228
5292 bp
lacO
T7 Promoter
NcoI (70)
BamHI (107)
EcoRI (113)
PstI (136)
HindIII (144)
lacO
T7 promoter
RBS
PstI (518)
FjTAL
ClaI (1271)
AvaI (1866)
T7 Terminator
aadA
ApaLI (2705)
RBS
Replication Origin
CloDF13 ori
lacI
ApaLI (4818)

Figure 7

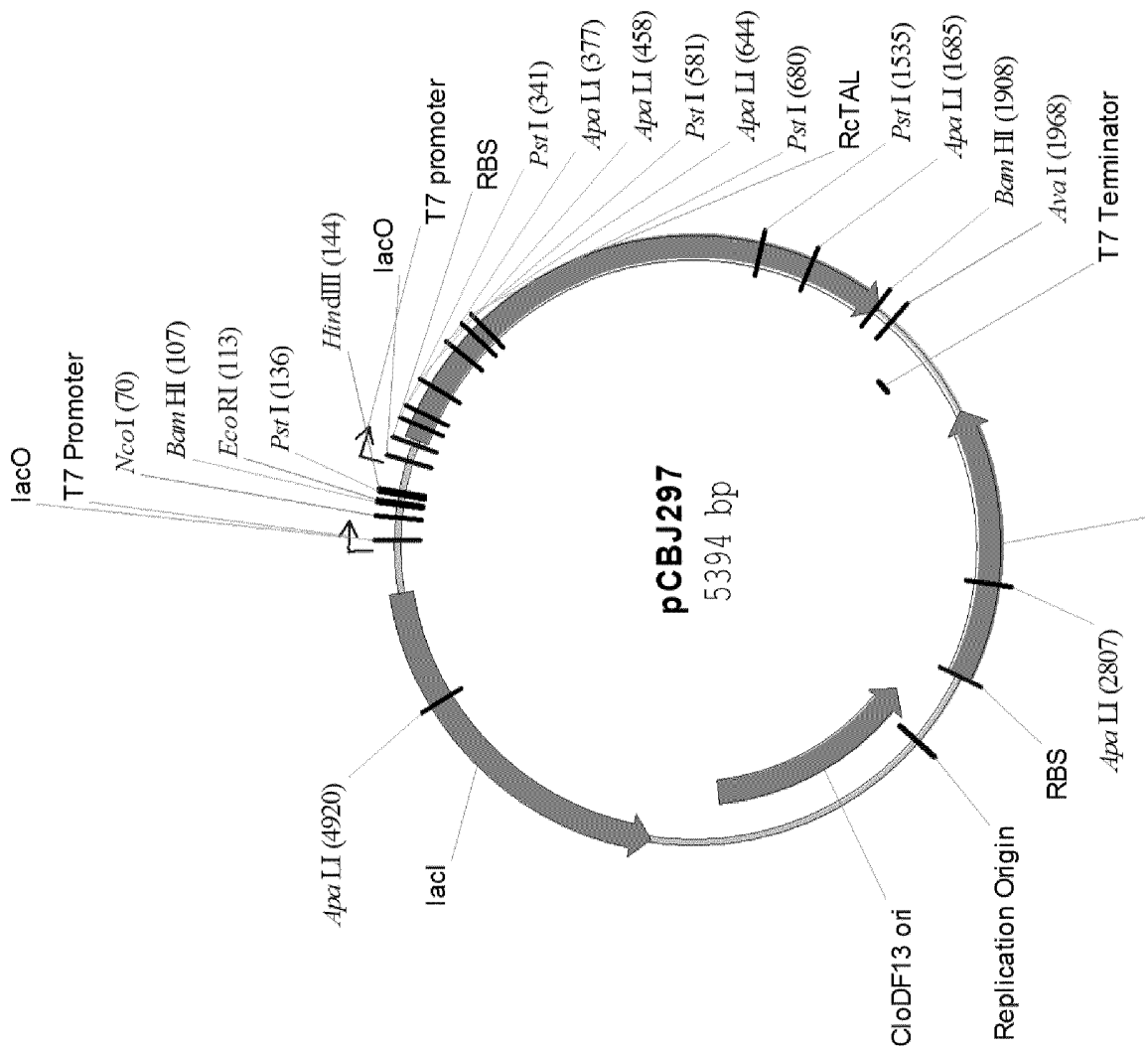

pCBJ297
5394 bp
lacO
T7 Promoter
NcoI (70)
BamHI (107)
EcoRI (113)
PstI (136)
HindIII (144)
lacO
T7 promoter
RBS
PstI (341)
ApaLI (377)
ApaLI (458)
PstI (581)
ApaLI (644)
PstI (680)
RcTAL
PstI (1535)
ApaLI (1685)
BamHI (1908)
AvaI (1968)
T7 Terminator
aadA
ApaLI (2807)
RBS
Replication Origin
CloDF13 ori
lacI
ApaLI (4920)

Figure 8

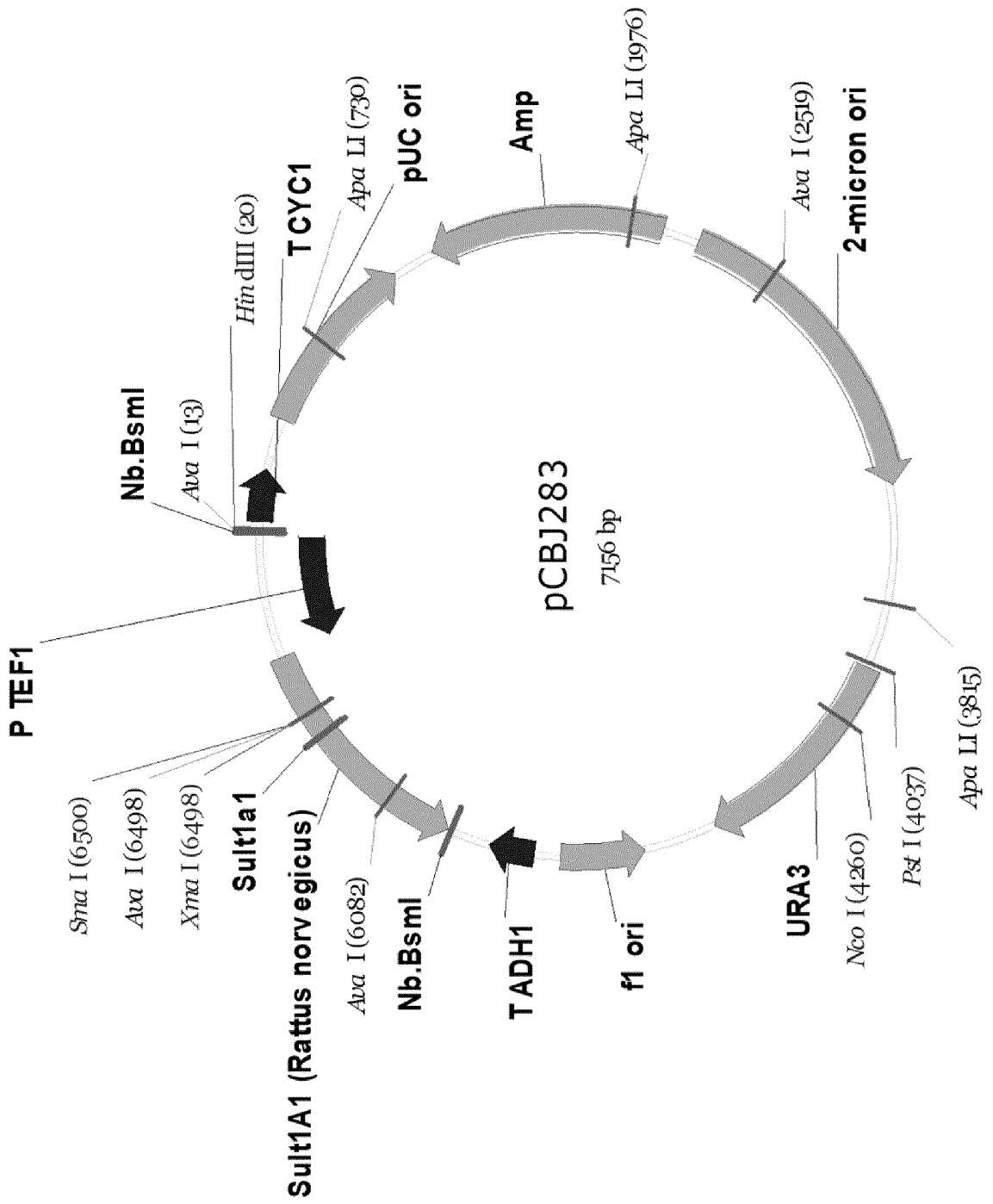

pCBJ283
7156 bp
P TEF1
Nb.BsmI
Ava I (13)
Hin dIII (20)
T CYC1
Apa LI (730)
pUC ori
Amp
Apa LI (1976)
Ava I (2519)
2-micron ori
Apa LI (3815)
Pst I (4037)
Nco I (4260)
URA3
f1 ori
T ADH1
Nb.BsmI
Ava I (6082)
Sult1A1 (Rattus norvegicus)
Sult1a1
Xma I (6498)
Ava I (6498)
Sma I (6500)

Figure 9

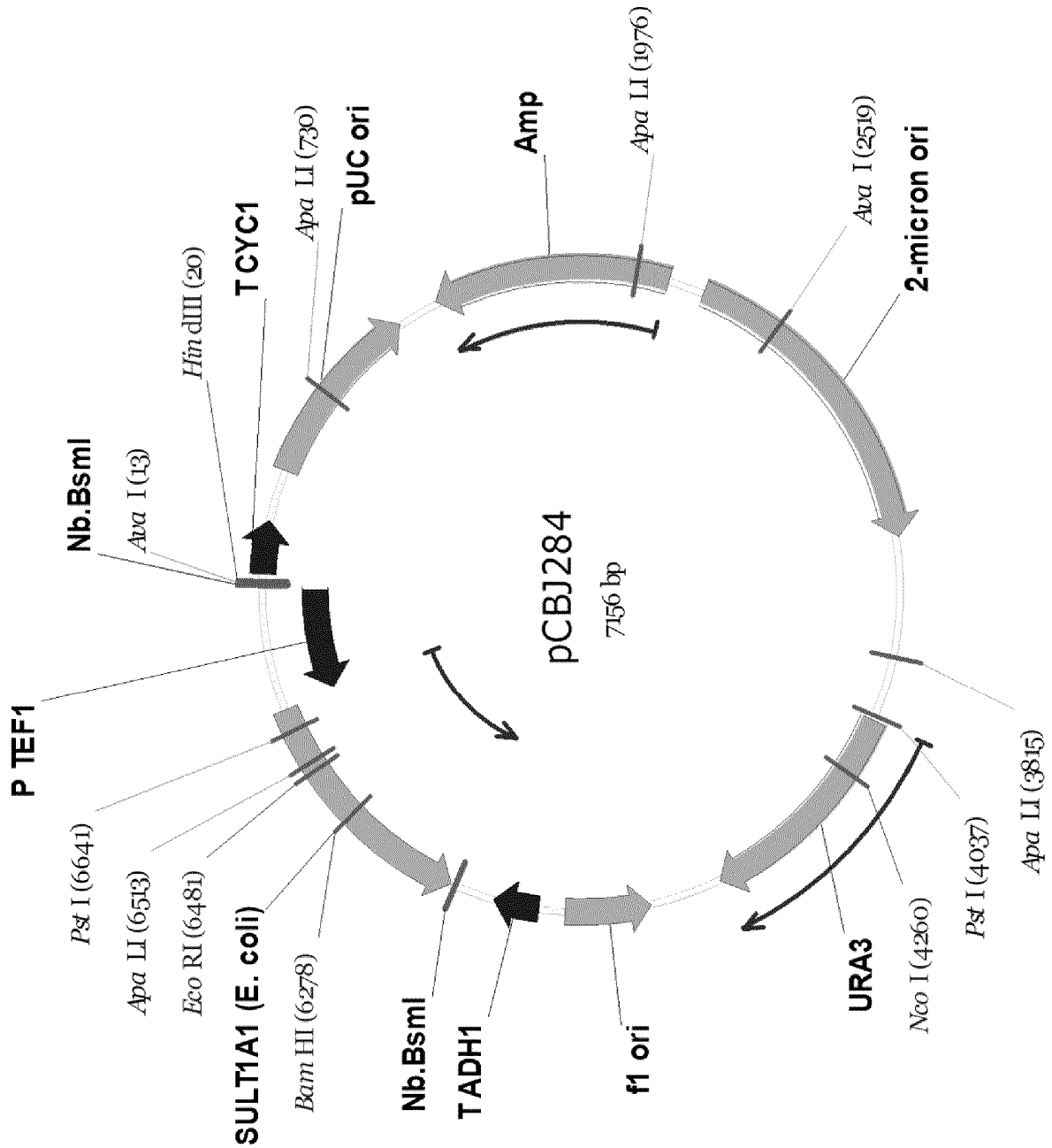

pCBJ284
7156 bp
P TEF1
Nb.BsmI
Ava I (13)
Hin dIII (20)
T CYC1
Apa LI (730)
pUC ori
Amp
Apa LI (1976)
Ava I (2519)
2-micron ori
Apa LI (3815)
Pst I (4037)
Nco I (4260)
URA3
f1 ori
T ADH1
Nb.BsmI
Bam HI (6278)
SULT1A1 (E. coli)
Eco RI (6481)
Apa LI (6513)
Pst I (6641)

Figure 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2015/069298

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/069298

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P11/00 C12N9/10
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12P C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, FSTA, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WONG ET AL: "Inhibition of hydroxycinnamic acid sulfation by flavonoids and their conjugated metabolites", BIOFACTORS, vol. 39, 2013, pages 644-651, XP002734418, * See page 644 (Abstract) *	1-38
Y	KAWAI ET AL: "p-Hydroxycinnamic acid production directly from cellulose using endoglucanase- and tyrosine ammonia lyase-expressing Streptomyces lividans", MICROBIAL CELL FACTORIES, vol. 12, 2013, pages 1-9, XP021151039, * See page 1 (Abstract) and page 2 (left column/lines 10-21) *	1-38

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 October 2015	19/10/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Korsner, Sven-Erik

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	AYUSO-FERNÁNDEZ ET AL: "Aryl sulfotransferase from Haliangium ochraceum: A versatile tool for the sulfation of small molecules", CHEMCATCHEM, vol. 6, 26 March 2014 (2014-03-26), pages 1059-1065, XP002734370, * See pages 1059-1060 (Introduction); online publication *	1-38
A	PURCHARTOVÁ ET AL: "Enzymatic preparation of silybin phase II metabolites: sulfation using aryl sulfotransferase from rat liver", APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, vol. 97, 2013, pages 10391-10398, XP035328725, * See page 10395 (Scheme 2) *	1-38
A	WONG ET AL: "In vitro and in vivo conjugation of dietary hydroxycinnamic acids by UDP-glucuronosyltransferases and sulfotransferases in humans", JOURNAL OF NUTRITIONAL BIOCHEMISTRY, vol. 21, 2010, pages 1060-1068, XP027432038, * See page 1063 (Table 2) *	1-38
A	BERGER ET AL: "The molecular basis for the broad substrate specificity of human sulfotransferase 1A1", PLOSONE, vol. 6, 2011, pages 1-10, XP002734371, * See page 1 (Introduction) *	1-38
A	PRATHER ET AL: "Golgi-resident PAP-specific 3'-phosphatase-coupled sulfotransferase assays", ANALYTICAL BIOCHEMISTRY, vol. 423, 2011, pages 86-92, XP028466644, * See page 86 (Abstract/last sentence) and page 87 (Figure 1); compare the Application, page 33, lines 22-24 *	1-38
Y	US 6 841 718 B2 (ALBERTE ET AL) 11 January 2005 (2005-01-11) * See Figures 1-8 and comments in column 3 *	16

-/--

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	JENDRESEN ET AL: "Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and Saccharomyces cerevisiae", APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 81, 24 April 2015 (2015-04-24), pages 4458-4476, XP055214411, * See page 4458 (Abstract); online publication *	12-15, 30-32, 35-38
Y,P	POYRAZ ET AL: "Crystal structures of the kinase domain of the sulfate-activating complex in Mycobacterium tuberculosis", PLOS ONE, vol. 10, 25 March 2015 (2015-03-25), pages 1-19, XP002745408, * See pages 1-2 (Abstract and Introduction) and page 3 (Figure 1); online publication *	6-9, 26-29
Y,P	VAN DER HORST ET AL: "Enzymatic sulfation of phenolic hydroxy groups of various plant metabolites by an arylsulfotransferase", EUROPEAN JOURNAL OF ORGANIC CHEMISTRY, vol. 2015, 12 December 2014 (2014-12-12), pages 534-541, XP002745414, * See pages 534-535 (Introduction) and page 537 (Figure 1); online publication *	1-38
A,P	HIRSCHMANN ET AL: "The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions", FRONTIERS IN PLANT SCIENCE, vol. 5, 16 October 2014 (2014-10-16), pages 1-13, XP002745392, * See page 1 (Abstract) and page 9 (from "How to identify...") - page 11; online publication *	1-38

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2015/069298

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6841718 B2	11-01-2005	AU 5975401 A	20-11-2001
		US 2002016980 A1	07-02-2002
		US 2006053510 A1	09-03-2006
		WO 0185971 A2	15-11-2001